INSTITUTO POTOSINO DE INVESTIGACIÓN CIENTÍFICA Y TECNOLÓGICA, A.C.

POSGRADO EN GEOCIENCIAS APLICADAS

Detection of deep convection clouds in quasi-real time using Machine Learning techniques and GOES-R data

Tesis que presenta

Rodrigo Dávila Ortiz

Para obtener el grado de

Doctor en Geociencias Aplicadas

Codirectores de la Tesis:

Dr. José Tuxpan Vargas

Dr. José Noel Carbajal Pérez

San Luis Potosí, S.L.P., 07 de junio de 2024

IPICYT

Constancia de aprobación de la tesis

La tesis "**Detection of deep convection clouds in quasi-real time using Machine Learning techniques and GOES-R data**" presentada para obtener el Grado de Doctor en Geociencias Aplicadas fue elaborada por **Rodrigo Dávila Ortiz** y aprobada el **07 de junio de 2024** por los suscritos, designados por el Colegio de Profesores de la División de Geociencias Aplicadas del Instituto Potosino de Investigación Científica y Tecnológica, A.C.

Dr. José Tuxpan Vargas
Codirector de la tesis

Dr. José Noel Carbajal Pérez
Codirector de la tesis

Dr. José Alfredo Ramos Leal
Miembro del Comité Tutoral

Dr. Juan Alberto Velázquez Zapata
Miembro del Comité Tutoral

Dr. José Luis González Compeán
Miembro del Comité Tutoral

Dr. Sócrates Figueroa Miranda
Miembro del Comité Tutoral

IPICYT

Créditos Institucionales

Esta tesis fue elaborada en el Laboratorio de Geomática y Modelación Numérica de la División Geociencias Aplicadas del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la codirección de los doctores José Tuxpan Vargas y José Noel Carbajal Pérez.

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología 841667 y del Instituto Potosino de Investigación Científica y Tecnológica, A. C.

Página en Blanco que se va a utilizar para colocar la copia del acta de examen.

A mis padres
María del Carmen Ortiz Rivera
Isaías Dávila Ramírez

Agradecimientos

A mis padres Carmen e Isaías, ustedes fueron mis primeros maestros y a quien les debo quien soy actualmente, espero honrar su recuerdo con este y todos los logros que sigan.

A Santy Dávila por su amor y apoyo, eres la razón que me motiva a ser mejor día con día.

A Héctor Dávila (Toro) por estar siempre ahí.

Al Dr. José Tuxpan, por su guía y asesoría durante la Maestría y el Doctorado. Su guía meticulosa y sus sugerencias constructivas fueron fundamentales para dar forma a este trabajo.

Al Dr. José Noel Carbajal, por su dirección durante la consolidación de este trabajo y su paciencia al explicarme los procesos físicos detrás de la formación de nubes.

Al Dr. Juan Alberto Velázquez, por su orientación invaluable, apoyo constante y sabios consejos desde la licenciatura a la fecha, y en especial por ser un gran amigo.

A los doctores, José Luis González Compeán, José Alfredo Ramos Leal y Sócrates Figueroa Miranda, por su participación activa en mi comité. Cada uno de ustedes aportó una perspectiva única y valiosa que enriqueció la calidad y la profundidad de esta investigación.

A mi familia Ruth, Gaby, Reny, Nacho y Juanpi por su cariño y apoyo. También a Mosty, Viejito, Christy, Dogo, Coffee y Panther.

A mis amigos, Paola, Marco, Dany, Pam y Raúl, son geniales chicos, gracias por todo.

A mis compañeros del Laboratorio de Geomática y Modelación Numérica, quienes más que compañeros los considero mis amigos, gracias por sus consejos, retroalimentación y apoyo.

Gracias al IPICYT, a la División de Geociencias Aplicadas y al Departamento de Control Escolar por su apoyo, seguimiento y brindarme siempre un espacio de aprendizaje, es un honor para mí formar parte de esta institución.

Gracias a todos los que de alguna manera contribuyeron a este logro. Este trabajo no habría sido posible sin su colaboración y respaldo.

Content

Constancia de aprobación de la tesis ii
Créditos institucionales iii
Acta de examen iv
Dedicatorias v
Agradecimientos vi
List of tables ix
List of figures x
Annexes xvi
Abbreviations xvii
Resumen xix
Abstract xx

Chapter 1. Introduction 21
1.1. Background 21
1.2. Problem Statement 22
1.3. Research Question 23
1.4. Justification 23
1.5. Objectives 24
1.5.1. General objective 24
1.5.2. Specific objectives 24
1.6. Test sites 24
1.6.1. Los Mochis, Sinaloa 24
1.6.2. Mexico City 26

Chapter 2. State of the art 28
2.1. Convective hazards assessment 28
2.2. The use of Machine and Deep Learning 30
2.3. Convective hazards monitoring and Early Warning Systems 32

Chapter 3. Proposed comprehensive framework 36
3.1. Phase 1: Characterization of storm clouds 37
3.2. Phase 2: Fusion of multi-source information 38
3.3. Phase 3: Hydrometeorological risk forecast 40

Chapter 4. Data and Methods 42
4.1. Data 42
4.1.1. ABI-GOES Data 42
4.1.2. GLM-GOES Data 43
4.1.3. GOES-R Products 44
4.1.4. MODIS Data 47
4.1.5. PERSIANN Data 47
4.2. Methodology: Deep convective events identification system 48

4.2.1. Preparation of Training Dataset 48
4.2.2. Machine Learning Approaches 50
4.2.3. Hyperparameters Tuning 57
4.2.4. Feature Scaling 59
4.2.5. Machine Learning Process 60
4.2.6. Workflow and implementation 61
4.2.7. Post-processing Lighting Filter 64
4.2.8. Accuracy metrics 65
4.3. Methodology: Fusion of multi-source information 67
4.3.1. Dimensionless precipitation curves 67
4.3.2. Extraction of deep convective cloud properties 70
4.3.3. Precipitation scenarios generation 71
4.4. Methodology: River flooding risk assessment 73
4.4.1. Runoff scenarios generation 73

Chapter 5. Results and Discussion 77
5.1. Feature importance analysis 77
5.2. Hyper-parameters Tuning results 81
5.3. Performance and validation of detection models 84
5.4. Simulation of reference convective events 91
5.5. Computational costs of ML approaches 96
5.6. Dimensionless precipitation curves for Los Mochis, Sin. 98
5.7. Generation of Rainfall-Runoff scenario assessment 99

Chapter 6. Example of application: Buoyancy force analysist 103
6.1. Deep convective cloud formation events 103
6.2. Deep convective clouds properties 106
6.3. Evolution of vertical development clouds (Velocity, Acceleration and Energy) 118

Chapter 7. Conclusions 122

References 125

Annexes 139

List of tables

Table 1. Specifications of ABI-GOES Bands. 42
Table 2. Specifications of ABI-GOES derived products. 44
Table 3. Overview of the input data for the implementation of ML workflow and subsequent post-processing filter. The initial choice of predictor variables (Interest Fields) was adopted from the work by S. Lee et al. (2017). 48
Table 4. ML models and approaches compared in this study. 51
Table 5. Hyperparameters of the LR, RF, and MLP tuned using the cross-validation strategy. 58
Table 6. Catchment parameter for RH10Fb basin. Source: SIATL-INEGI (2016). 76
Table 7. Interest Fields were used as predictor variables after the feature importance analysis, for Los Mochis and Mexico City datasets. 78
Table 8. Summary of the hyperparameters of LR, RF, and, MLP were tuned using the cross-validation strategy. 83
Table 9. IoU metric results for the Los Mochis 2018-227 - 18:00 (August 15, 2018), Los Mochis 2019-234 - 17:30 (August 22, 2019), Mexico City 2019-247 - 20:10 (September 04, 2019), and Mexico City 2021-135 - 19:50 (May 15, 2021) events for each ML Approach. 91
Table 10. List of events included in this example of application. 103

List of figures

Figure 1. Land surface elevation over Los Mochis. The red border corresponds to the spatial simulation domain, which has a size of 50 x 50 cells. Source: INEGI (2017).	25
Figure 2. Cumulative annual precipitation during the period 1964-2012 estimated at the Mochis - 25116 climate station (CLICOM, 2016). Figure adapted from Dávila Ortiz et al. (in press).	26
Figure 3. Land surface elevation over the study area called Mexico City. The red border corresponds to the spatial simulation domain, which has a size of 50 x 50 cells. Source: INEGI (2017).	27
Figure 4. Example of a classification of different convective weather hazards and their relationships to thunderstorms. Source: (Hayward et al., 2020).	28
Figure 5. The “Early Warnings for All” initiative pillars. Source: https://wmo.int/news/media-centre/early-warnings-all-initiative-scaled-action-ground.	34
Figure 6. Early Warning Systems implemented in Mexico. Source: Centro Nacional de Prevención de Desastres (2023).	35
Figure 7. Proposed methodological framework.	36
Figure 8. GOES-R CONUS domains for Operational GOES-West Location (Green), Check-out Location (Red) and Operational GOES-East Location (Blue). Source: Lindstrom (2017).	43
Figure 9. GOES R series instruments. Source: https://cimss.ssec.wisc.edu/education/goesr/goesr1.html?version=mobile.	43
Figure 10. GOES-16 ABI imagery for each of the instrument’s 16 bands (CONUS) on December 18, 2017. Source: https://www.goes-r.gov/products/baseline-cloud-moisture-imagery.html	45
Figure 11. ISCCP Cloud Classification in terms of Cloud Top Pressure and Cloud Optical Depth. Source: Rossow & Schiffer (1991).	49
Figure 12. Distribution of Los Mochis, Sin and Mexico City datasets. For Los Mochis dataset, the number of classes labeled as CC is 24,116 and noCC is 10,884, whereas for the Mexico City dataset, the number of classes labeled as CC is 57,833 and noCC is 79,667.	50
Figure 13. Diagram of ensemble ML methods. a) Bagging, b) Voting, c) Stacking.	51
Figure 14. Architecture of a Logistic Regression (left; Torres et al., 2019) and Random Forest (right; Sahour et al., 2021).	52
Figure 15. An example of the general-type decision tree used by Gagne et al. (2009) to classify convective areas using DTs (left), and the attribute sets used (right).	53
Figure 16. Definition of the “margin” between classes: the criterion that SVM try to optimize. Source: https://la.mathworks.com/discovery/support-vector-machine.html	54
Figure 17. Example of MLP Network adapted for the CC and noCC classification task.	55

Figure 18. Comparison between (a) grid search and (b) random search for hyperparameter tuning. The nine dots represent candidates. The curves on the left and top denote the model accuracy (*e.g.*, NMSE or overall accuracy) as a function of each search dimension. Source: Salgado Pilario et al. (2021). 58

Figure 19. In cross-validation, the dataset is divided into k equally sized smaller subsets, called folds. A model is trained using k-1 of the folds as training data, leaving one for testing in each iteration, providing a robust evaluation of a machine learning model's performance. Source: https://scikit-learn.org/stable/modules/cross_validation.html. 59

Figure 20. Schematization of the StandardScaler method. Standardize features by removing the mean and scaling to unit variance. Source: https://app.dataquest.io/m/744/k-means-with-scikit-learn-and-interpreting-results/3/scaling-the-data. 60

Figure 21. General scheme of the framework for the identification and monitoring of potential convective events in quasi-real time (5 min). The ML approaches used in the ML process were selected based on their performance parameters during the training and testing phases at each study site. 62

Figure 22. Schematization of the Lightning Filter generation process. When the presence of a Lightning Event is identified from the GOES GLM data accumulated in 5 min within the site of interest, it is associated with a cell of the model special domain, using its distance from the nearest cell center as the criterion for assignment. The Lightning Cells (LC), together with a buffer of one cell per edge (BC), are called Convection Cells (CC) and are integrated into the output matrix of the ML models. 64

Figure 23. A confusion matrix is a table used in machine learning to evaluate the performance of a classification algorithm. It provides a summary of the model's predictions compared with the actual results, grouping the results into four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Source: https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826. 66

Figure 24. Schematic of the approach used to identify and separate storms (a). Storm mass curves for the identified storms (b). Dimensionless storm mass curves (c). Intersections of storm mass curves with vertical lines (d). Source: Bonta (2004). 68

Figure 25. Storm mass−curve intersections with isopleths of probability (Huff curves) and NRCS design storms ("Type" curves) superimposed. Source: Bonta (2004). 69

Figure 26. Schematic of the deep convective cloud property extraction process. The CC pixels are detected (a) and extraction masks are generated from them in the array stack of the GOES R products; TPW, CtH, CtP, CtT, and COD (b). The values were averaged, and time series were constructed to represent the behavior of these variables during the evolution of the potential deep convection event detected. 71

Figure 27. Los Mochis study site, the red boundary represents the ABI-GOES domain and the grid corresponds to the spatial domain of the PERSIANN database (0.25° x 0.25°). The yellow dots correspond to the PERSIANN cells used to generate the dimensionless precipitation curves, and the red dot represents the location of Los Mochis-25116 weather station (CLICOM, 2016). 72
Figure 28. Daily precipitation records between January 2000 and February 2013 for Los Mochis, test site. Source: Created with information from (CLICOM, 2016). 72
Figure 29. Coordinates of the Dimensionless Unit Hydrograph. Source: Wanielista & Yousef (1992). 74
Figure 30. RH10Fb basin (green) and ABI-GOES simulation domain for Los Mochis test site (red). 75
Figure 31. Feature importance analysis using the absolute value of the coefficients obtained in the training stage of the LR model (a-b) and the MDI estimated for the 12 Interest Fields with the RF model (c-d). a-c corresponds to Los Mochis and b-d to Mexico City datasets. A high MDI value and the weight of the coefficients are interpreted as indicators of the degree of contribution of a variable. 77
Figure 32. Box plots of input variables for reference data used in the CC detection models for Los Mochis. The predictive dataset was generated from the most significant features of the LR and RF models. The number of classes labeled as CC is 24,116 and noCC is 10,884. 80
Figure 33. Box plots of input variables for reference data used in the CC detection models for Mexico City. The predictive dataset was generated from the most significant features of the LR and RF models. The number of classes labeled as CC is 57,833 and noCC is 79,667. 80
Figure 34. Hyperparameter tuning results for RF (a, b), LR (c, d), and MLP (e, f). The plots on the left correspond to Los Mochis dataset (a, c, e), and those on the right correspond to the Mexico City dataset (b, d, f). 82
Figure 35. Evaluation results of the POD, FAR, ACC, BIAS, CSI, and F1 metrics for the ML approaches using Los Mochis test dataset. Each vertex of the graph corresponds to an ML approach, the blue lines correspond to the performance results of the ML models (ML), and the red lines show the results with Lightning Filter integration (ML+LF). 85
Figure 36. Evaluation results of the POD, FAR, ACC, BIAS, CSI, and F1 metrics for the ML approaches using Mexico City test dataset. Each vertex of the graph corresponds to an ML approach, the blue lines correspond to the performance results of the ML models (ML), and the red lines show the results with Lightning Filter integration (ML+LF). 86
Figure 37. ROC curves and AUC values for the predictions obtained with LR (black), RF (red), MLP (brown), LR-Stacking (blue), RF-Stacking (cyan), Bagging (magenta), and Soft-Voting (light green) for Los Mochis (a) and Mexico City (b) datasets. These ROC curves provide a comparison of the sensitivity and specificity of each ML approach at different discrimination thresholds. 91

Figure 38. IoU maps of the Los Mochis 2018-227 - 18:00 (August 15, 2018) event, generated with the reference labels and simulations of the eight ML approaches LR (a), RF (b), MLP (c), Bagging (d), LRstacking (e), RFstacking (f), Hvoting (g), Svoting (h), and Cloud-top brightness temperature map of the same event obtained from the ABI-GOES channel 14 (i). In general, most ML models underestimate the cloud boundary, except for LR, where the identification of a potential unlabeled convective core is simulated. The remnants of this core are also presented in MLP, Bagging, and Svoting. Salt-and-pepper noise is observed in LRstacking and RFstacking. 92

Figure 39. IoU maps of the Los Mochis 2019-234 - 17:30 (August 22, 2019), event generated with the reference labels and simulations of the eight ML approaches LR (a), RF (b), MLP (c), Bagging (d), LRstacking (e), RFstacking (f), Hvoting (g), Svoting (h), and Cloud-top brightness temperature map of the same event obtained from the ABI-GOES channel 14 (i). The majority of the ML models showed the tendency to underestimate the CC class at the edge of the convective cloud and the presence of salt-and-pepper noise in MLP, Bagging, LRstacking and RFstacking and Svoting 93

Figure 40. IoU maps of the Mexico City 2019-247 - 20:10 (September 04, 2019), event generated with the reference labels and simulations of the eight ML approaches LR (a), RF (b), MLP (c), Bagging (d), LRstacking (e), RFstacking (f), Hvoting (g), Svoting (h), and Cloud-top brightness temperature map of the same event obtained from the ABI-GOES channel 14 (i). The outputs of LR, RF, MLP, Bagging, and Hvoting are consistent with each other and show a significant level of overestimation at the edge of the cloud. On the other hand, the outputs generated by LRstacking, RFstacking, and Svoting are extremely noisy, especially at the boundary of the clouds. 94

Figure 41. IoU maps of the Mexico City 2021-135 - 19:50 (May 15, 2021), event generated with the reference labels and simulations of the eight ML approaches LR (a), RF (b), MLP (c), Bagging (d), LRstacking (e), RFstacking (f), Hvoting (g), Svoting (h), and Cloud-top brightness temperature map of the same event obtained from the ABI-GOES channel 14 (i). The simulations obtained with LRstacking, RFstacking, and Svoting are particularly noisy; however, the remaining ML approaches correctly simulate deep convection zones with a significant degree of underestimation at the edges of these regions. 95

Figure 42. Computational cost graph for each ML approach corresponding to Los Mochis (a) and Mexico City (b). The right axis shows the total simulation time and the left axis shows the memory usage. 97

Figure 43. Step1: Storm mass curves for storms identified. Storm totals >40 mm (a). Step2: Dimensionless storm mass curves (b). Step3: Storm mass−curve intersections with Duration times (c). Step4: Isopleths of probability (Huff Curves; d). 98

Figure 44. Comparison between RF and LR simulations with Cloud-top brightness temperature map for Los Mochis 2018-261 18:57 event (Tropical Depression 19-E). 100
Figure 45. a) Estimated average TPW values using detected CC pixels as the extraction mask array for each satellite view of the 19-E event over the Los Mochis domain (each 5 min). b) 50% Isopleths of probability. 101
Figure 46. Precipitation scenario derived from the average TPW of simulated CC pixels in Los Mochis 2018-261 18:57 event (Tropical Depression 19-E) and the 50% Huff isopleth. The hyetograph has a duration of 6 h and a total precipitation depth of 58 mm. 102
Figure 47. Estimated hydrograph (Runoff scenario) for the RH10Fb basin, based on the precipitation scenario generated for Los Mochis 2018-261 18:57 event (Tropical Depression 19-E). 102
Figure 48. Selected spatial domains in this example of application. a) Trans-Mexican Volcanic Belt section, b) Chiapas-Tabasco, c) Los Mochis, Sin, d) Ciudad Victoria, Tamps, and e) Álvarez Mountain Range. 104
Figure 49. Binary labeling of a convective event using a thresholding approach for different time steps for Event 01 (Trans-Mexican Volcanic Belt section). Yellow cells correspond to CC and purple cells to no CC. 105
Figure 50. Relief representation of the spatial domains in 3D. The red border indicates state boundaries, while the black border indicates coastlines. a) Trans-Mexican Volcanic Belt section, b) Chiapas-Tabasco, c) Los Mochis, Sin, and d) Ciudad Victoria, Tamps. The 3D models were generated using information from INEGI (2017). 106
Figure 51. Temporal behavior of Cloud-top Height (CtH) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps). 107
Figure 52. Schematization of type of clouds. Source: https://www.aopa.org/news-and-media/all-news/2016/august/flight-training/weather. 108
Figure 53. 3D Representation of the Cloud-top Height (CtH) for Event 02 (Trans-Mexican Volcanic Belt section), Event 05 (Chiapas-Tabasco), Event 07 (Los Mochis, Sin), and Event 09 (Ciudad Victoria, Tamps). 109
Figure 54. Diagram of a typical deep convective cloud. Source: https://weathertogether.net/weather-101/convective-clouds/ 110
Figure 55. Temporal behavior of Cloud-top Pressure (CtP) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps). 111
Figure 56. Temporal behavior of Cloud-top Temperature (CtT) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps). 112
Figure 57. Cloud-top brightness temperature maps for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps). 113
Figure 58. Temporal behavior of Cloud Optical Depth (COD) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas- 115

Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps).

Figure 59. Mean Cloud Optical Depth (COD) maps for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps). 116

Figure 60. Temporal behavior of Total Precipitable Water (TPW) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps). 117

Figure 61. Figure 61. Rate and Acceleration of Convection for the Event 02 (a, e; Trans-Mexican Volcanic Belt section), Event 05 (b, f; Chiapas-Tabasco), Event 07 (c, g; Los Mochis, Sin), and Event 09 (d, h; Ciudad Victoria, Tamps). 121

Annexes

Annex 1. Dávila-Ortiz, R., Carbajal-Pérez, J. N., Velázquez-Zapata, J. A., & Tuxpan-Vargas, J. (2024). Approximation of a Convective-Event-Monitoring System Using GOES-R Data and Ensemble ML Models. *Remote Sensing*, 16(4), 675. 140

Annex 2. Graphical Abstract of the paper: Approximation of a Convective-Event-Monitoring System Using GOES-R Data and Ensemble ML Models. 167

Annex 2. Dávila Ortiz, R., Tuxpan Vargas, J., & Velázquez Zapata, J. A. (2023). Identification of Deep Convection Clouds Using ABIGOES Data and Machine Learning Techniques: The Case of Los Mochis, Sinaloa, Mexico. 2023 Mexican International Conference on Computer Science (ENC), 1–7. https://doi.org/10.1109/ENC60556.2023.10508667 168

Annex 3. Dávila Ortiz, R., Tuxpan Vargas, J., & Velázquez Zapata, J. A. (in press). La producción del riesgo ante eventos de inundación en Los Mochis, Sinaloa. In Riesgos y desastres relacionados con agua transformación del territorio, inundaciones y contaminación. El Colegio de San Luis, A.C. 175

Annex 4. Dávila Ortiz, R., & Tuxpan Vargas, J. (2024, enero 11). Cazando nubes de tormenta. Propuesta para un sistema de monitoreo y alerta temprana basado en inteligencia artificial en México. Opinión Pública SLP. https://opslp.mx/cazando-nubes-de-tormenta-propuesta-para-un-sistema-de-monitoreo-y-alerta-temprana-basado-en-inteligencia-artificial-en-mexico/ 206

Abbreviations

ABI	Advanced Baseline Imager
AHI	Advanced Himawari Imager
AUC	Area Under Curve ROC
AWS	Amazon Web Services
CAPE	Convective Available Potential Energy
CC	Convective Cells
CENAPRED	Centro Nacional de Prevención de Desastres
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station
CHRS	Center for Hydrometeorology and Remote Sensing
CI	Convective Initiation
CMIP	Cloud and Moisture Imagery
CNN	Convolutional Neural Networks
CO	Convective Occurrence
COD	Cloud Optical Depth
COMS	Communication, Ocean, and Meteorological Satellite
CONAGUA	Comisión Nacional del Agua
CONUS	Continental United States
COT	Cloud Optical Thickness
CP	Cloud Phase
CPSD	Cloud Particle Size Distribution
CSI	Critical Success Index
CSM	Clear Sky Mask
CtH	Cloud-top Height
CtP	Cloud-top Pressure
CtT	Cloud-top Temperature
DL	Deep Learning
DMW	Derived Motion Winds
DNN	Depth Neural Networks
DSI	Derived Stability Indices
DT	Decision Tree
DUH	Dimensionless Unit Hydrograph
ENSO	El Niño–Southern Oscillation
EWS	Early Warning System
FAR	False alarm ratio
GLM	Geostationary Lightning Mapper
GOES	Geostationary Operational Environmental Satellite
IDF	Intensity-Duration-Frequency

INEGI	Instituto Nacional De Estadística y Geografía
INIFAP	Instituto Nacional de Investigaciones Forestales y Agropecuarias
IoU	Intersection over Union
ISCCP	International Satellite Cloud Climatology Project
LF	Lightning Filter
LR	Logistic Regression
MDI	Mean Decrease Impurity
MDPD	Minimum Dry Period Duration
ML	Machine Learning
MLP	Multi-layer Perceptron
MODIS	Moderate-Resolution Imaging Spectroradiometer
NAM	North American Monsoon
noCC	Non-Convective Cells
NWP	Numerical weather predictions
OT	Overshooting Top
PCA	Principal Component Analysis
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
POD	Probability of detection
RF	Random Forest
RNN	Recurrent Neural Networks
ROC	Receiver Operating Characteristic
SCS	Soil Conservation Service
SGD	Stochastic Gradient Descent
SMN	Servicio Meteorológico Nacional
SMO	Sierra Madre Occidental
SVM	Support Vector Machine
TO	Thunderstorm Occurrence
TPW	Total Precipitable Water
WMO	World Meteorological Organization

Resumen

Detección de nubes de convección profunda en tiempo cuasi-real basado en técnicas de Aprendizaje de Maquina y datos GOES-R.

La presencia de nubes de convección profunda está directamente relacionada con potenciales peligros convectivos como incidencia de rayos, caída de granizo, tormentas severas, inundaciones repentinas, tornados, etc. Por otro lado, México cuenta con una red limitada y heterogénea de instrumentos para el monitoreo y pronóstico eficiente y confiable de dichos eventos. En este estudio, se desarrolló un sistema para la identificación y modelación de nubes de convección profunda en tiempo cuasi-real utilizando modelos de Aprendizaje Automático. Este sistema se basa en tres principios clave: el uso de software de código abierto y productos geoespaciales de libre acceso, la automatización y escalabilidad a todo el territorio mexicano. En total, ocho diferentes modelos de Aprendizaje Automático: Regresión Logística (LR), Bosque Aleatorio (RF), Perceptrón Multicapa (MLP), Apilamiento con LR y RF, Embolsado, y Votación Suave y Dura fueron entrenados utilizando Campos de Interés derivados a partir de datos del sensor Advanced Baseline Imager (ABI) a bordo del Satélite Ambiental Operativo Geoestacionario - Serie R (GOES-R). Estos modelos fueron evaluados en dos sitios de estudio: Los Mochis, Sinaloa y la Ciudad de México, seleccionados por su intensa actividad convectiva y alta vulnerabilidad ante eventos climáticos extremos. Los hallazgos demuestran que un enfoque simple como LR o RF puede identificar y simular eficazmente nubes de convección profunda en ambas áreas de estudio, logrando una Probabilidad de Detección (POD) ≈ 0.84 para Los Mochis y POD ≈ 0.72 para la Ciudad de México. Además, se obtuvieron valores del índice de Falsas alarmas (FAR) de aproximadamente 0.2 para Los Mochis y 0.4 para la Ciudad de México. Posteriormente, un filtro de post-procesamiento basado en la incidencia de rayos (Lightning Filter) utilizando datos del Geostationary Lightning Mapper, carga útil del satélite GOES-16, mostró un potencial significativo para mejorar los valores de POD registrados por los modelos de Aprendizaje Automático. Finalmente, el sistema de detección de nubes de convección profunda se integró con datos de precipitación multi-fuente para generar escenarios de precipitación en tiempo cuasi-real y realizar evaluaciones de riesgo ante tormentas severas y análisis de avenidas fluviales. Este estudio es pionero en el desarrollo e implementación de un Sistema de Alerta Temprana para los peligros asociados con actividad convectiva en México.

PALABRAS CLAVE: Aprendizaje automático; datos ABI-GOES; previsión de peligros convectivos; nubes de convección profunda; sistemas de alerta temprana; sistema en tiempo cuasi-real.

Abstract

Detection of deep convection clouds in quasi-real time using Machine Learning techniques and GOES-R data

The presence of deep convective clouds is directly related to potential convective hazards such as lightning, hail, severe storms, flash floods, and tornadoes. On the other hand, Mexico has a limited and heterogeneous network of instruments for efficient and reliable monitoring and forecasting of such events. In this study, a quasi-real-time framework to identify and model deep convective clouds using Machine Learning (ML) models was developed. This framework was guided by three key principles: the use of open-source software and open-access geospatial products, automation, and scalability.

Eight different ML models and model ensemble approaches, including Logistic Regression (LR), Random Forest (RF), Multi-Layer Perceptron, LRstacking, RFstacking, Bagging, Hard-Voting, and Soft-Voting, were trained using Interest Fields derived from Advanced Baseline Imager (ABI) sensor data aboard the Geostationary Operational Environmental Satellite - R Series (GOES-R). These models were evaluated at two study sites: Los Mochis and Mexico City, which were selected for their intense convective activity and high vulnerability to extreme weather events. The findings demonstrate that a simple approach such as LR or RF can effectively identify and simulate deep convective clouds in both study areas, achieving a probability of detection (POD) ≈ 0.84 for Los Mochis and POD ≈ 0.72 for Mexico City. In addition, false alarm ratio (FAR) values of approximately 0.2 for Los Mochis and 0.4 for Mexico City were obtained. Subsequently, a post-processing filter based on lightning incidence (Lightning Filter) using data from the Geostationary Lightning Mapper (GLM) onboard the GOES-16 satellite showed significant potential to improve the POD of the ML models. Furthermore, the deep convective cloud detection framework was integrated with multi-source precipitation data to generate quasi-real-time precipitation scenarios and perform risk assessments for severe storms and river flooding. This study sets a precedent towards the development and implementation of an Early Warning System for hazards associated with intense convective activity in Mexico.

KEY WORDS: Machine learning; ABI-GOES data; convective hazard forecasting; deep convective clouds; early warning systems; quasi-real-time framework.

Chapter 1. Introduction

1.1. Background

Convective hazards, which are characterized by dynamic atmospheric processes involving the vertical movement of air masses, pose significant challenges to human safety and infrastructure resilience. These hazards include a wide spectrum of high-impact phenomena such as thunderstorms, wind, hail, tornadoes, lightning, and flooding (Cancelada et al., 2020; H. Han et al., 2015; S. Lee et al., 2017; Liu et al., 2019; Pereira-Nunes et al., 2024).

According to the WMO Atlas of Mortality and Economic Losses from Weather, Climate, and Water Extremes (World Meteorological Organization, 2021), 11,072 disasters related to weather, climate, and water extremes were recorded globally between 1970 and 2019. Of these, 44% were attributed to floods and 35% to severe storm events. In addition, extreme storms accounted for 39% of reported deaths, whereas floods contributed to 16% (2,064,929 total reported deaths), with both types of disasters causing the most significant economic losses. Other types of convective hazards, such as lightning, hail, or tornadoes, also contribute to disasters with human and economic losses on a global scale (World Meteorological Organization, 2021).

Due to its physical and geographical characteristics, Mexico presents a high degree of vulnerability to extreme hydrometeorological events (Conde et al., 2016). Furthermore, various sectors of the population, influenced by socioeconomic conditions, are particularly susceptible to the effects of extreme convective events. The WMO Atlas of Mortality and Economic Losses reports 6,655 deaths from 202 weather-related disasters in Mexico, with storms and floods being the most frequent (World Meteorological Organization, 2021).

Given the rapid evolution and microscale variability of deep convection events, forecasting, and implementing early mitigation measures becomes inherently complex. Moreover, many countries, including Mexico, lack reliable observation networks and short-range weather forecasting systems (León-Cruz et al., 2023). The proposed framework in this study presents an alternative for implementing an Early Warning System (EWS) in areas highly vulnerable to convective hazards and with limited meteorological monitoring infrastructure.

This work will focus primarily on small convective cells rather than more significant phenomena such as squall lines and mesoscale convective complexes. (*i.e.*, larger convective systems that can persist for several hours and cover a large region). This is because small convective cells can form quickly and last only a short time; however, they have the potential to produce localized thunderstorms and heavy rain (Pereira-Nunes et al., 2024).

1.2. Problem Statement

Deep convective clouds are directly related to potential convective hazards such as lightning strikes, hail, severe storms, flash floods, and tornadoes. On the other hand, Mexico has a limited and heterogeneous network of instruments that allow efficient and reliable monitoring and forecasting of such events. Leveraging Machine Learning (ML) techniques and the high-resolution data provided by the GOES-R series of satellites presents an opportunity to enhance the monitoring and forecasting of deep convective clouds. Despite the potential benefits, there remains a need for comprehensive research into the proposal and implementation of a deep convective clouds monitoring system that effectively integrates ML algorithms with multi-source information. This need is particularly pronounced in regions such as Mexico, which are characterized by complex atmospheric dynamics, limited severe weather monitoring infrastructure, and a highly variable distribution of ground-based sensors.

1.3. Research Question

How can ML techniques and the integration of multi-source information be effectively used to develop, implement, and scale a deep convective cloud monitoring system for Mexico, and what are the critical considerations for its successful deployment to improve severe weather forecasting and EWS?

1.4. Justification

The dissertation proposal addresses a critical gap in the Mexican territory's current meteorological monitoring and forecasting capabilities by focusing on developing and implementing an innovative deep convective cloud monitoring system. By using ML algorithms and taking advantage of the high-resolution data provided by the GOES-R satellite constellation, the proposed system can significantly improve the detection, tracking, and forecasting of severe weather phenomena associated with deep convective clouds and the estimation of precipitation scenarios in quasi-real-time (every 5 min). Integrating ML techniques with multi-source information, including GOES-R data, provides improved spatial and temporal resolution opportunities, enabling more accurate and timely severe weather forecasts.

Furthermore, the research conducted in this study will contribute to the advancement of both geospatial ML applications and the generation of tools for studying deep convection events in Mexico. This will have broader implications for the prevention and mitigation of risks associated with storm events and flood occurrence by providing timely and accurate information to decision-makers and vulnerable populations.

By addressing this research question, this dissertation aims to set a precedent for implementing an EWS for hazards associated with intense convective activity in Mexico. This region has complex atmospheric dynamics, limited severe weather monitoring infrastructure, and a highly heterogeneous distribution of ground-based sensors.

1.5. Objectives

1.5.1. General objective

Develop a new alternative, affordable, automated, and scalable storm cloud monitoring, and EWS based on ML techniques, continuous processing of GOES-R data, and feedback with multi-source climatological data.

1.5.2. Specific objectives

- To develop a deep convective cloud identification framework based on three principles: open-source software and open-access products, automation, and scalability.

- To compare different ML approaches to determine the optimal performance and computational cost, using Los Mochis, Sinaloa, and Mexico City as case studies.

- To integrate the deep convective cloud detection framework with multi-source precipitation data to generate quasi-real-time precipitation scenarios.

- To establish a methodological basis for monitoring and EWS of floods associated with storm clouds by integrating precipitation scenarios with hydrological modeling techniques.

- To apply the framework for detecting potential deep convection events to the study and analysis of storm cloud formation processes and their physical properties in different regions of Mexico.

1.6. Test Sites

1.6.1. Los Mochis, Sinaloa

Los Mochis, the main city of the Ahome municipality, is located in the northeastern region of Sinaloa, Mexico. It falls within the North American monsoon (NAM) core domain, which is characterized by significant convective activity (Cavazos et al., 2008; Mejia & Douglas, 2009; Ramos-Pérez et al., 2022) and a substantial increase in summer rainfall (Forzieri et al., 2011). The distribution of precipitation in this area is primarily influenced by the deep convective cloud pattern, as noted by Ramírez-López et al. (2023). Figure 1 depicts the location of Los Mochis, which is between the Gulf of California to the west and the Sierra Madre Occidental (SMO) to the east. The simulation area, comprising a grid of 50 x 50 ABI-GOES cells, lies within the latitudinal range of 25.25° to 26.5° and the longitudinal range of -110° to -108. The topographic configuration in this region is primarily flat, with mountainous systems at the edge of the SMO. This region is classified as Hot Desert climate (BWh; INEGI, 2008), with a mean annual precipitation of 335 mm, primarily concentrated during a pronounced wet season spanning from July to September. The contrast between the Pacific Ocean and the Gulf of California waters and the complex topography of the two mountain ranges suggests that they are agents of spatiotemporal variability of humid atmospheric processes (Ramírez-López et al., 2023).

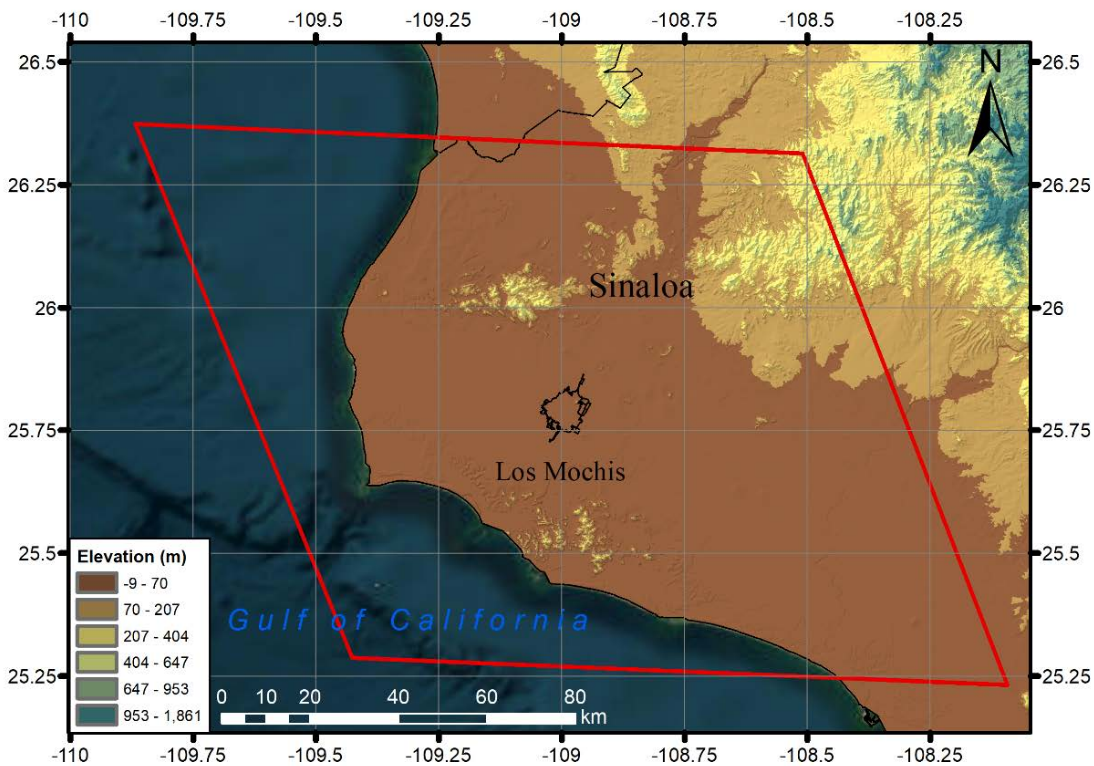


Figure 1. Land surface elevation over Los Mochis. The red border corresponds to the special simulation domain, which has a size of 50 x 50 cells. Source: INEGI (2017).

Precipitation patterns are significantly influenced by large-scale climatic phenomena such as El Niño–Southern Oscillation (ENSO) and hurricanes (Dávila Ortiz, 2019). For example, in 1984, when El Niño occurred, the accumulated annual precipitation was 683.7 mm (Figure 2); in 1996, when Hurricane Faust occurred, the precipitation was 599.3 mm (Figure 2); another particularly rainy year was 2008, when Hurricane Olaf occurred, and the precipitation was 595.4 mm (584 mm according to IMPLAN, 2012).

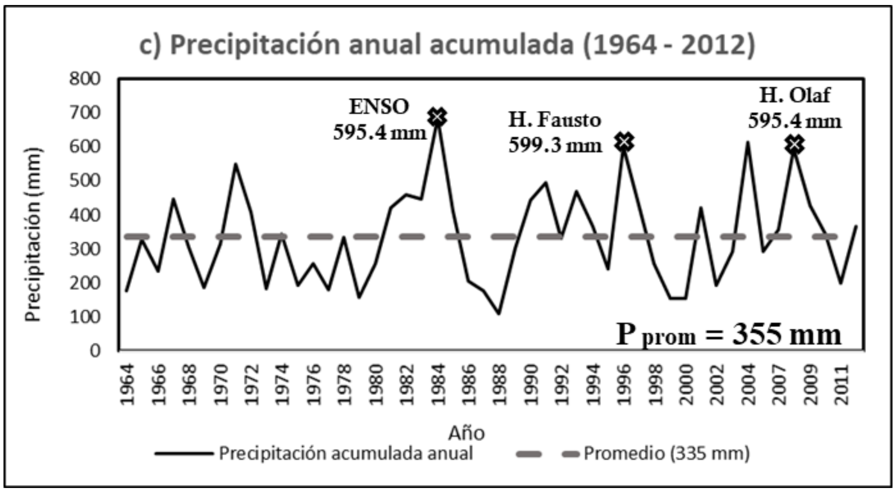


Figure 2. Cumulative annual precipitation during the period 1964-2012 estimated at the Mochis - 25116 climate station (CLICOM, 2016). Figure adapted from Dávila Ortiz et al.(2023b).

1.6.2. Mexico City

Mexico City, the capital and largest city of Mexico, is the major population center in the country. Situated in central Mexico, it lies within the physiographic province known as the Trans-Mexican Volcanic Belt, which crosses the country from the Pacific Ocean to the Gulf of Mexico. This region presents a complex topography, with elevations ranging from sea level to 5,000 meters above sea level (masl). The combination of different factors, such as its latitudinal position, complex topography, the influence of tropical cyclones, cold fronts, and easterly waves (León-Cruz et al., 2021), and even the effect of urbanization (Vargas & Magaña, 2020), cause this area to present intense convective activity. Consequently, these conditions are a source of convective hazards associated with tornadoes (Carbajal et al., 2019; León-Cruz et al., 2021), floods (Romero Lankao, 2010), hail, and thunderstorms (León-Cruz et al., 2023). In addition, socioeconomic factors in the area contribute to a high level of social vulnerability (Eakin et al., 2016).

Figure 3 shows the geographic location of the Mexico City area (the spatial domain of the simulation includes the states of México and Morelos), which comprises a grid of 50 x 50 ABI-GOES cells and lies within the northern latitude range of 20° to 19° and the longitude range of -99.75° to -98.5°. The mean annual precipitation in Mexico City ranges from 621 to 1,200 mm (León-Cruz et al., 2021), and this zone presents a Subhumid Temperate climate type (Cw; INEGI, 2008)

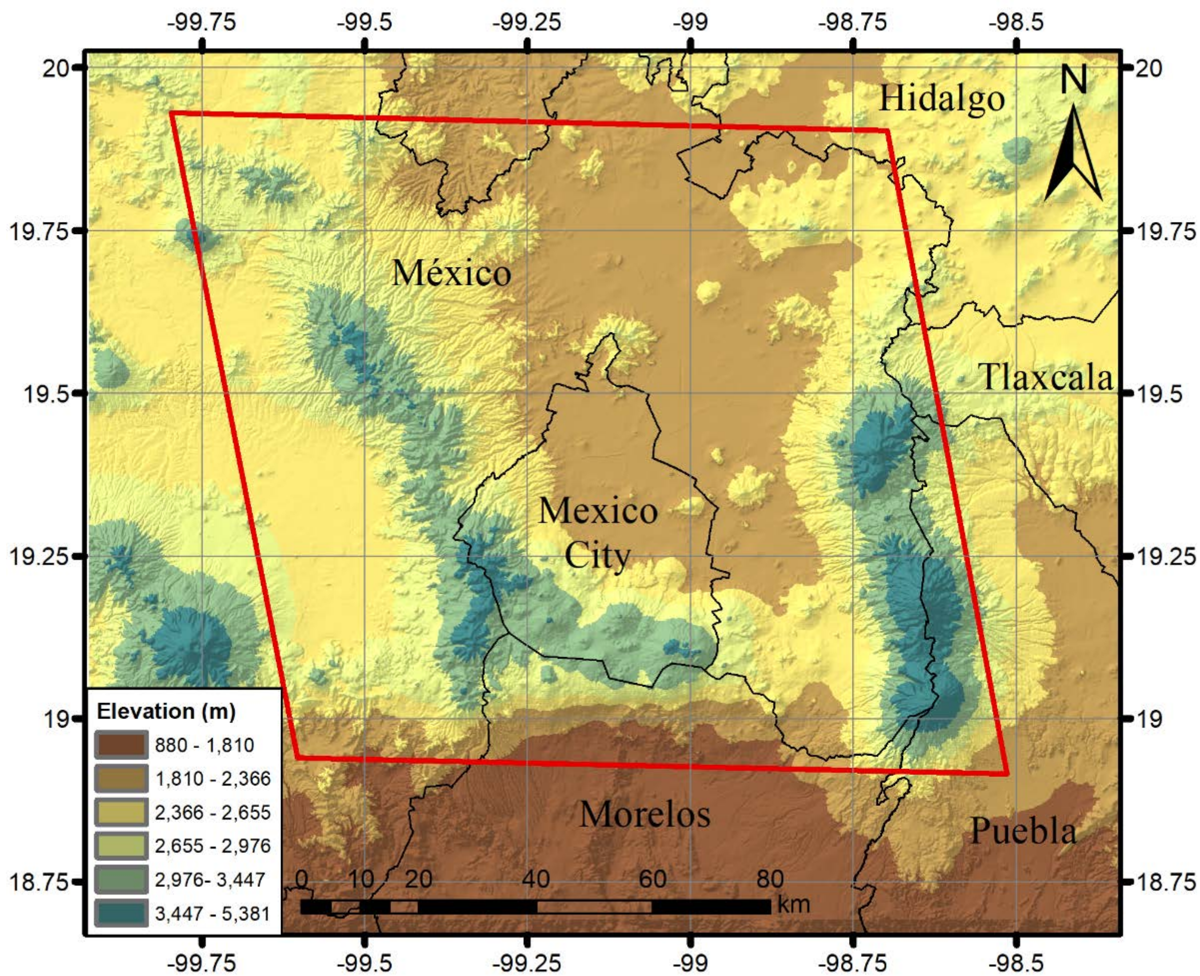


Figure 3. Land surface elevation over the study area called Mexico City. The red border corresponds to the spatial simulation domain, which has a size of 50 x 50 cells. Source: INEGI (2017).

Chapter 2. State of the art

2.1. Convective hazards assessment.

High-impact convective weather events, such as thunderstorms, tornadoes, hailstorms, lightning, and downbursts (Figure 4), result in significant fatalities, economic losses, and infrastructure damage (McGovern et al., 2023). Given the profound societal impacts of these atmospheric phenomena, it is crucial to study them.

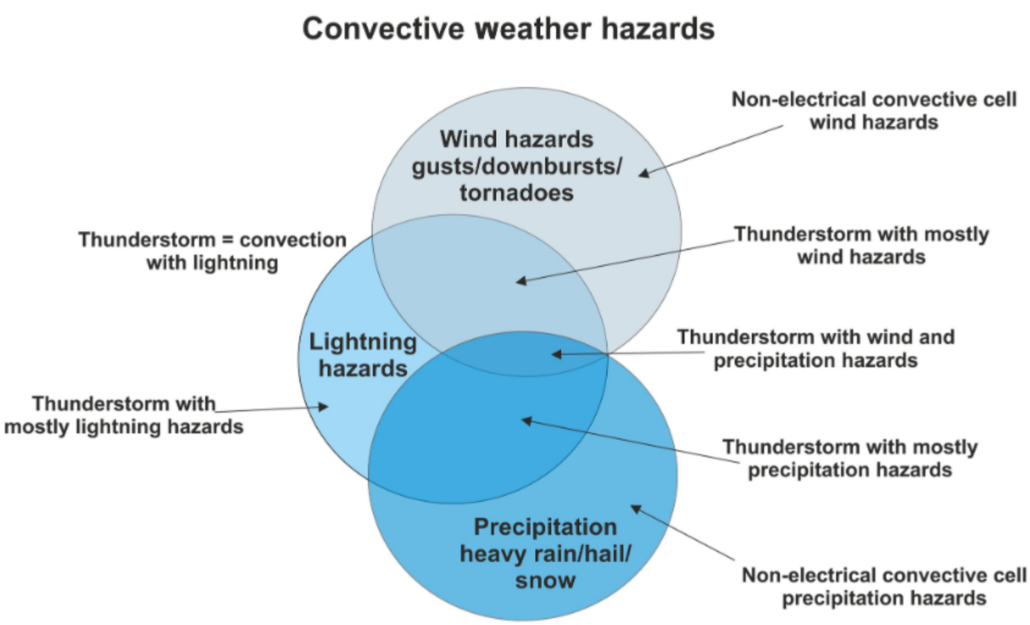


Figure 4. Example of a classification of different convective weather hazards and their relationships to thunderstorms. Source: Hayward et al. (2020).

Despite the critical importance of studying convective hazards, several limitations and challenges exist within this research field. One significant challenge arises from the complex interactions between various atmospheric factors, including temperature, humidity, wind patterns, air pressure, and their connections to large-scale climate signals. For instance, Tippett et al. (2015) identified close relationships between hazardous convective weather and climate signals such as the Madden–Julian Oscillation (MJO), ENSO, and radiatively forced climate change, as evidenced by studies analyzing tornado probability (Barrett & Gensini, 2013), hail occurrence (Barrett & Henley, 2015), and convective storms (Brooks, 2013). Understanding these processes requires sophisticated mathematical models and observational data, which may have inherent limitations and uncertainties.

Another major limitation in the study of this type of hazard is its spatial and temporal variability and limited predictability. Convective hazards can exhibit significant spatial and temporal variability, making it difficult to accurately predict their occurrence, intensity, and precise location. This variability can result from local topography (*e.g.*, Carbajal et al., 2019; León-Cruz et al., 2021), land cover (*e.g.*, Tracy & Mecikalski, 2023), and other factors that influence atmospheric dynamics (Tracy & Mecikalski, 2023). Similarly, despite advances in numerical weather prediction models, there are inherent limits to the predictability of convective hazards, especially at longer lead times. Factors such as chaotic behavior, small-scale variability, and uncertainties in initial conditions can limit forecast accuracy (Prein et al., 2021).

Finally, high-quality observational data are essential for studying convective hazards and developing accurate forecast models. However, there may be limitations in data availability, especially in remote or sparsely monitored, and data quality issues such as gaps, biases, and inaccuracies. Along these lines, Tippett et al. (2015) highlighted the need for more high-quality observations of convective weather for hazards, with no foreseeable solution to this problem. In addition, current observational collection practices still need to be improved.

Convective hazard assessment is an emerging field of research, despite the inherent challenges and limitations. Several studies have focused on developing algorithms for Convective Initiation (CI) nowcasts, *i.e.*, these short-term weather forecasts predict the initiation and development of deep moist convective activity (Weckwerth & Parsons, 2006).

CI nowcasts rely on various observational data sources and methods, for example, radar imagery (*e.g.*, the Thunderstorm Identification, Tracking, Analysis, and Nowcasting; TITAN; Dixon & Wiener, 1993), atmospheric instability indices (*e.g.*, Mueller et al., 1993), satellite data (Cancelada et al., 2020; Cintineo et al., 2020; Gravelle et al., 2016; Roberts & Rutledge, 2003), numerical weather prediction (NWP, *e.g.*, the Corridor Integrated Weather System; CIWS; Veillette, 2013), Weather Research & Forecasting (WRF) Model outputs (*e.g.,* Pereira-Nunes et al., 2024) and other meteorological parameters to identify favorable conditions for thunderstorm initiation.

For instance, the Satellite Convection Analysis and Tracking (SATCAST; Mecikalski & Bedka, 2006) is a CI nowcasting expert system that uses eight predictors called "Interest Fields" based on infrared Geostationary Operational Environmental Satellite (GOES) data to forecast CI with 0-1 h lead times (McGovern et al., 2023). In this work, CI is defined as the first detection of Weather Surveillance Radar-1988 Doppler (WSR-88D) reflectivities ≥ 35 dBZ produced by convective clouds and satellite-derived atmospheric motion vectors (AMVs) for tracking individual cumulus clouds. Subsequently, a second upgrade of this system, SATCAST version 2 (STACASTv2), was proposed by Walker et al. (2012) and includes object tracking approaches to compute temporal changes in brightness temperature. Later additions include the integration of NWP data (Veillette, 2013).

2.2. The use of Machine and Deep Learning

In recent years, integrating ML techniques into the geosciences, facilitated by advances in satellite data quality (Reichstein et al., 2019), has spurred various research efforts focused on CI identification. For example, Mecikalski et al. (2015) proposed a CI algorithm that combines Interest Fields derived from GOES-R and NWP model data, using Logistic Regression (LR) and Random Forest (RF) models to achieve better probabilistic predictions compared with binary approaches. Through experiments, they validated the performance of these CI algorithms in the United States, with a FAR of 0. 1-0.18 lower than existing deterministic CI detection algorithms for GOES (FAR ≈ 0.48 - 0.6; Walker et al., 2012).

In the related literature, numerous studies have reported the development and implementation of CI nowcast algorithms based on ML approaches for various geostationary satellites, such as Himawari-8 (D. Han et al., 2019; S. Lee et al., 2017), the Communication, Ocean, Meteorological Satellite (COMS; H. Han et al., 2015), and Meteosat (Krinitskiy et al., 2023), and GOES (Mecikalski & Bedka, 2006).

Other satellite data-based approaches developed to aid in the diagnosis of preconvective environmental characteristics and deep moist convection are the Convective Occurrence (CO) algorithms, which include approaches such as modeling convective initiation/decay and storm advection (McGovern et al., 2023), the 0-9 h NearCast model (Gravelle et al., 2016; Petersen et al., 2010), and object tracking techniques such as optical flow (Lenk et al., 2018) and thunderstorm occurrence (TO; La Fata et al., 2021; Ukkonen & Mäkelä, 2019; Zhou et al., 2020).

Overshooting cloud Top (OT) detection algorithms are another common type of deep convective detection approach used to identify and analyze cloud formations, focusing on overshooting cloud tops. These formations (also called anvil domes) are defined as dome-like clouds that form over a cumulonimbus cloud top or penetrate the tropopause (American Meteorological Society, 2023). The OT clouds form when strong updrafts in thunderstorms push cloud tops above their equilibrium height (level of neutral buoyancy). These OT cloud tops are often associated with severe weather events. Detection methods for OT clouds are typically based on cloud top temperature thresholds (*e.g.*, K. M. Bedka et al., 2018; Khlopenkov et al., 2021), as well as ML and Deep Learning (DL) approaches (*e.g.*, Kim et al., 2017, 2018).

Regarding DL applications, this field has recently gained traction in studying convective weather. For example, a pioneering study by Y. Lee, Kummerow, & Ebert-Uphoff (2021) implemented a convolutional neural network (CNN; LeCun et al., 1998) to detect ongoing convection, with the advantage that they used a sophisticated echo-classification algorithm to generate their labels, which uses richer radar information than a simple reflectivity threshold, the labeling method used in most studies (McGovern et al., 2023).

Similarly, Lagerquist et al. (2021) used DL techniques, specifically U-nets, to forecast CO at 0–2 h lead times. U-nets, a convolutional neural network architecture commonly used for semantic segmentation tasks, feature a U-shaped structure with contracting and expanding pathways to capture local features and global context in image data (Ronneberger et al., 2015). Like S. Lee et al. (2017) and D. Han et al. (2019) they used infrared Interest Fields from the Himawari-8 satellite. Other examples of this type of U-net architecture being addressed for high-impact weather hazard issues are Hilburn et al. (2020) and Samsi et al. (2019).

The integration of DL models significantly improves convective weather forecasting by capturing intricate spatial and temporal dynamics. This advancement enhances our ability to predict the formation, evolution, and intensity of convective storms with increased accuracy.

2.3. Convective hazards monitoring and Early Warning Systems

"An Early Warning System (EWS) is an integrated system that facilitates preparedness and response mechanisms through the dissemination of early warning to reduce the impact of a natural disaster" (Agbehadji et al., 2023).

The EWS are indispensable tools that can help save lives and reduce the impact of disasters on infrastructure, making it particularly important in areas of high vulnerability. For instance, it is estimated that USD 800 million is spent annually on developing and operating EWS in developing countries that lack the resources to mitigate the effects of natural disasters (WMO, 2022). However, in a global context, it is estimated that only one in three people can access early warning services (Agbehadji et al., 2023).

There is a wide variety of EWS to address various types of disasters, both natural and anthropogenic, such as earthquakes (Zheng et al., 2022), flash droughts (X. Yuan et al., 2019), air pollution (Jiang et al., 2019), debris flows, and tsunamis (Galaz et al., 2022). Technological advances have substantially enriched these systems in the context of convective hazards. Contemporary methodologies incorporate cutting-edge techniques such as artificial intelligence, fuzzy logic models, wavelet transform, acoustic emission techniques, dynamic predictors, real-time satellite observations, big data analytics, Internet of Things integration, and ML and DL algorithms (Agbehadji et al., 2023).

Examples of EWS for convective weather purposes include the study by Harjupa et al. (2022) to detect and predict the occurrence of heavy rainfall in Indonesia, Yucel & Onen (2014) a satellite-based algorithm to estimate extreme precipitation events in northwestern Turkey, Qing et al. (2022) a tornado EWS based on radar data, Mahomed et al. (2021) with their Ground-Based Lightning Detection and Near-Real-Time Warning System, among others.

W. Yuan et al. (2021) present a notable example of EWS implementation. Cumulative Distribution Functions (CDFs) were applied to fit the cumulative rainfall-duration curves corresponding to typical rainfall patterns and the Probability Density Functions (PDFs). Subsequently, the HEC-HMS hydrological model was applied to simulate the rainfall-runoff process.

Floods are the most frequent type of natural disaster, presenting significant challenges to flood EWS, particularly in the accurate detection of floods to prevent damage to property and lives (Wannachai et al., 2022), the determination of the warning module threshold, and the development and calibration of hydraulic models (Cools et al., 2012). Some illustrative frameworks for EWS include those proposed by Bartos et al. (2018); Cools et al. (2012); Ritter et al. (2020); Wannachai et al. (2022).

Globally, EWS is a critical element in achieving sustainable development and increasing resilience, especially in areas with high levels of marginalization. In this context, in November 2022, the United Nations (UN) Secretary-General presented the Executive Action Plan for the “Early Warning for All” initiative (WMO, 2022), to be implemented between 2023 and 2027, with an initial investment of approximately

US$ 3.1 billion, a minimal amount compared to the benefits that will be derived from the initiative, according to the WMO (United Nations, 2022).

The Early Warnings for All initiative is built on four pillars to deliver effective and inclusive multi-hazard early warning systems, disaster risk information, hazard monitoring and forecasting, warning dissemination and communication, and disaster response capability (Figure 5).

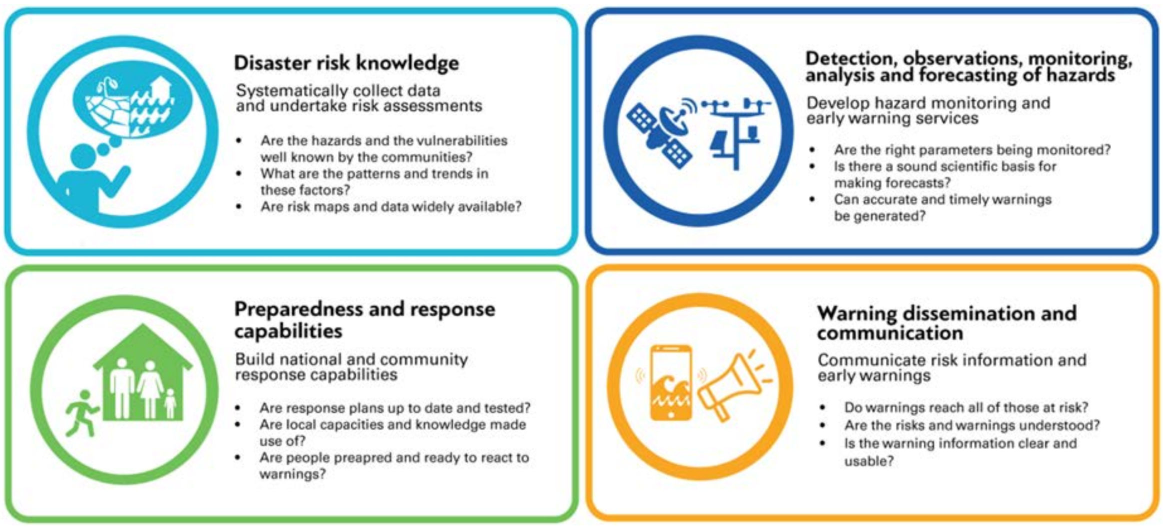


Figure 5. The “Early Warnings for All” initiative pillars. Source: https://wmo.int/news/media-centre/early-warnings-all-initiative-scaled-action-ground.

In Mexico, according to the Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT, 2017): Faced with this vulnerability, to save lives, Mexico has created seven EWS with different coverage and warning times (Figure 6): Servicio Meteorológico Nacional (National Meteorological Service) (1877), Servicio Sismológico Nacional (National Seismological Service) (1910), Sistema de Alerta Sísmica Mexicano (Mexican Seismic Alert System) (1991), de Sistema de Monitoreo del Popocatépetl (Popocatepetl Monitoring System) (1994), Sistema de Alerta Temprana para Ciclones Tropicales (Early Warning System for Tropical Cyclones) (2000) y Sistema Nacional de Alerta de Tsunamis (National Tsunami Warning System) (2013).

These EWS, managed by the Centro Nacional de Prevención de Desastres (National Disaster Prevention Center) (CENAPRED), are based on four main components:

1. Prior knowledge and identification of risks associated with disruptive phenomena.
2. A system for measuring and monitoring the disruptive phenomenon in order to make forecasts or issue scientifically based warnings.
3. Dissemination of public alerts with clear and precise information that activates the response of the population.
4. Response or contingency plans to know what to do when faced with the impact of disruptive phenomena.

EARLY WARNING SYSTEM
MÉXICO

SEGURIDAD
SECRETARÍA DE SEGURIDAD Y PROTECCIÓN CIUDADANA

CNPC
COORDINACIÓN NACIONAL DE PROTECCIÓN CIVIL

System	Phenomenon	Information	Coverage	Start date	Moment of Notice
Servicio Sismológico Nacional	Seismic	www.ssn.unam.mx	National	1910	Earthquake warning
Sistema de Alerta Sísmica Mexicano (SASMEX)	Seismic	www.cires.org.mx	Mexico City, Oaxaca, Chilpancingo, Acapulco and Morelia	1991	Seconds before the arrival of an earthquake that has already occurred. It depends on the distance from the epicenter and the energy of the earthquake.
Sistema de Monitoreo del Volcán Popocatépetl	Volcanic	www.Gob.mx/cenapred	Areas surrounding a volcano	1994	If an event occurs
Sistema de Alerta Temprana para Ciclones Tropicales	Tropical Cyclone	www.preparados.gob.mx	National	2000	Up to 72 hours in advance
Sistema Nacional de Alerta de Tsunamis	Tsunami	digaohm.semar.Gob.mx/cat/centroAlertasTsunamis.html	Pacific Coast and Gulf of Mexico	2013	For local tsunamis, minutes in advance, for regional and distant or transoceanic tsunamis, hours.
Sistema de Alerta Temprana de Incendios en México	Forest fires	www.gob.mx/conabio	National	1999	If an fire occurs
Servicio Meteorológico Nacional	Hydrometeorological	smn.conagua.gob.mx	National	1877	Notice if any event occurs and forecasts

Source: CNPC - CENAPRED

Figure 6. Early Warning Systems implemented in Mexico. Source: Centro Nacional de Prevención de Desastres (2023).

Despite global and national efforts to implement EWS for various disruptive phenomena, significant opportunities for research, innovation, and action remain. Incorporating artificial intelligence, machine learning, and big data analytics can substantially enhance the accuracy of risk prediction models and the speed of response to potential threats.

Chapter 3. Proposed comprehensive framework

The methodological framework presented in Figure 7 consists of an alternative for monitoring and EWS of floods associated with storm clouds by analyzing and continuously processing data from the GOES-R platform and feedback with multi-source climatological data. It is based on integrating three modules or phases with two main characteristics: open-access products and open-source software and the potential scalability to any area within the national territory.

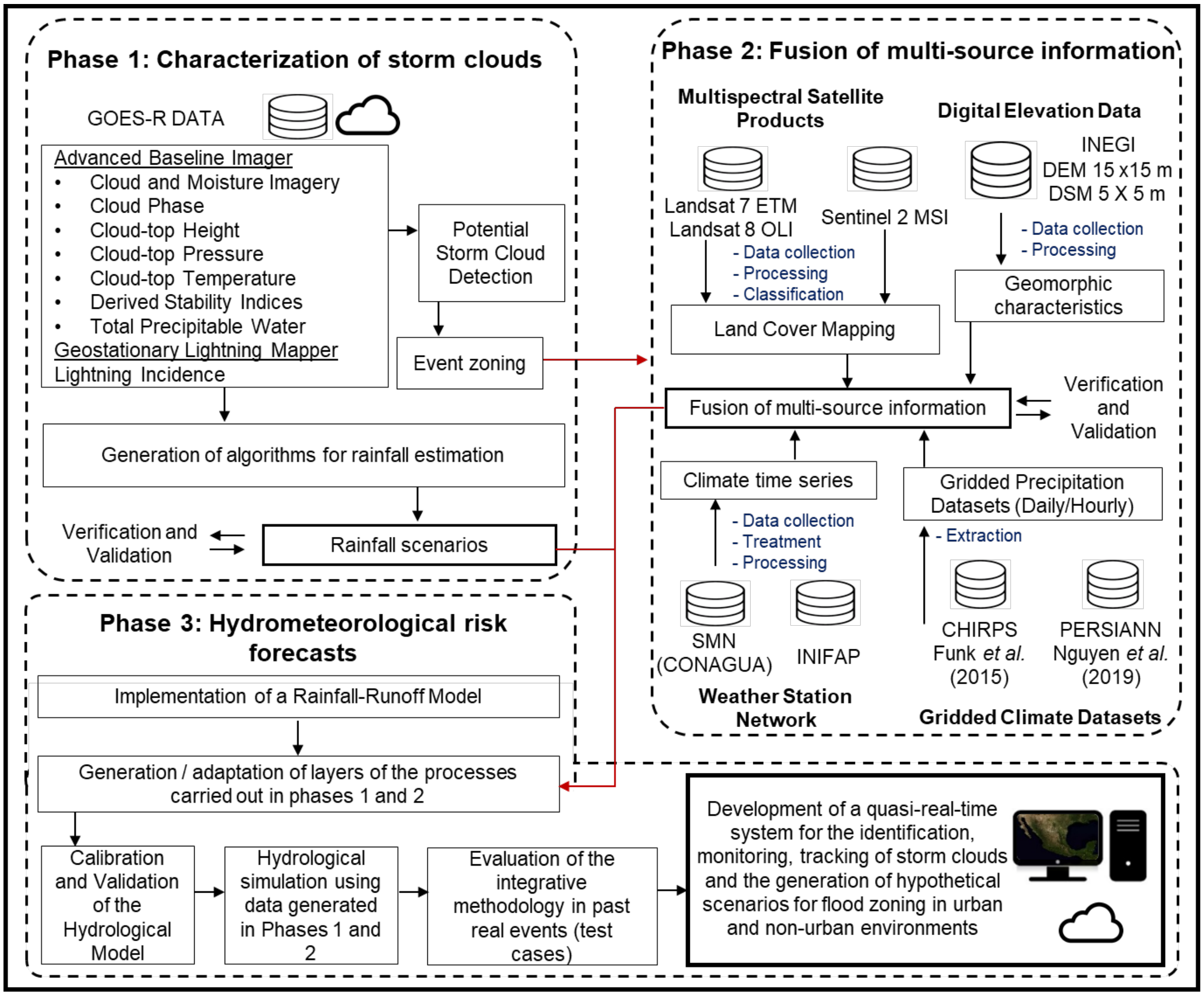


Figure 7. Proposed methodological framework.

These modules are as follows: Phase 1: Characterization of storm clouds, Phase 2: Fusion of multi-source information, and Phase 3: Hydrometeorological risk forecast. A general description of each phase is provided below.

A necessary clarification is that a fundamental tenet of the EWS frameworks is that the forecasting model includes a monitoring model, a communication strategy, and an emergency plan to help in managing natural disasters (Agbehadji et al., 2023; Calvello & Piciullo, 2016). This effort requires integrating interdisciplinary knowledge between scientists, researchers, and stakeholders, which is beyond the scope of the framework proposed here but is fertile ground for future research.

3.1. Phase 1: Characterization of storm clouds

It consists of an algorithm for identifying, monitoring, and tracking potential deep convective events (storm clouds) based on Machine Learning (ML) techniques and information obtained from the ABI sensor of the GOES-16 satellite, which is continuously updated every 5 minutes. Its implementation requires an exhaustive analysis of the performance of each ML method included in an ensemble of different models, which must be previously trained with a dataset of reference deep convection events. It is necessary to delimit one or more areas of interest. A detailed description of the methods, data, and steps proposed for implementing this system will be provided in later sections. Furthermore, the results of this phase have been reported by Dávila Ortiz et al. (2023) and Dávila-Ortiz et al. (2024).

From the ML models, a binary classification is made between a class called Convective Cells (CC), which corresponds to the presence of a deep convective cloud, and a second class called Non-Convective Cells (noCC), which corresponds to clear skies or any cloud cover that is not a vertically developing cloud. The result of this classification, generated every 5 min, allows tracking and monitoring of the evolution and displacement of this type of formation. The event's zoning is performed based on the spatial distribution of the detected CC pixels.

Finally, the pixels classified as CC serve as a mask for the extraction of various optical and physical properties of the storm clouds, including the height, pressure, and temperature at the top of the cloud, the optical depth of the cloud (*i.e.*, a measure of the transparency or thickness of clouds, indicating how much they attenuate or scatter incoming sunlight), and the Total Precipitable Water (TPW) content, which are derived from the different GOES-R products. In this sense, the information obtained from the ABI-GOES sensor can be divided into two groups. Spectral information was extracted from the 16 bands of the ABI-GOES sensor and the cloud products derived from them. In Chapter 4, these two types of products are described in detail.

The amount of TPW provides an estimate of precipitation for each of the pixels classified as CC. However, for simplicity, these values are averaged to provide a mean precipitation value for the simulated convective event at a given time. In the second phase of this framework, once a storm cloud is identified and the average amount of TPW is estimated, this value is fitted to a precipitation distribution constructed from an analysis of historical precipitation data for the simulation domain.

3.2. Phase 2: Fusion of multi-source information

This phase has two main objectives. The first step is to typify all the terrain characteristics related to the simulation area's hydrological and hydraulic behavior, such as land use and type of vegetation cover or soil classification. The second objective is to analyze the distribution of the precipitation regime of the area based on the climatic records collected by the meteorological station networks and the use of gridded precipitation datasets. From the precipitation data, rainfall distributions will be estimated to adjust the mean TPW values once a potential deep convection event is detected, thus generating precipitation scenarios. The primary sources of information considered for this phase are listed below.

- Multispectral Satellite Products: Landsat 7 Enhanced Thematic Mapper (ETM), Landsat 8 Operational Land Imager (OLI), and Sentinel 2 MultiSpectral Instrument (MSI) are used for land cover mapping using supervised classification techniques. Maps of land use and vegetation cover are used to estimate infiltration potential using the Horton (1933) or Green & Ampt (1911) methods, in addition to mapping the Manning roughness coefficient (*N*) by cover type (Chow, 1959). Information about the identification of the soil type can be extracted from the edaphological layers produced by the Instituto Nacional de Investigaciones Forestales y Agropecuarias (INIFAP & CONABIO, 2001). For more information on generating thematic layers from satellite imagery, see Dávila Ortiz (2019).

- Digital Elevation Data: The proposed source of information for the digital elevation models (DEMs) is the Mexican Elevation Continuum in its version 3.0 (CEM 3.0; INEGI, 2013) for a spatial resolution of 15 x 15 m and the set of high-resolution surface digital elevation models (5 x 5 m; INEGI, 2017).

- Climate time series: Information from the Climatic Database for Mexico CLICOM (2016) of the Servicio Meteorológico Nacional (SMN) was used to identify days with storms based on an analysis of their historical rainfall records. On the other hand, using information from the National Network of Automated Agrometeorological Stations of the INIFAP (2018) and the Network of Automated Meteorological Stations of the Comisión Nacional del Agua (CONAGUA) through the SMN (SMN & CONAGUA, 2023) rainfall intensity data were extracted for each storm and recorded daily.

- Gridded Precipitation Datasets: Gridded precipitation datasets address the spatio-temporal limitations of data availability in station networks within the simulation domain. Two spatially distributed gridded precipitation products are proposed, one with daily and one with hourly temporal resolution. The CHIRPS (Climate Hazards Group InfraRed Precipitation with Station; Funk et al., 2015) is a high-resolution global precipitation dataset combining satellite-derived infrared measurements with meteorological station data for accurate and timely precipitation information. The PERSIANN (Precipitation Estimation from Remotely Sensed Information using

Artificial Neural Networks; Nguyen et al., 2019) dataset is a global precipitation dataset derived from satellite observations and artificial neural network algorithms for improved precipitation monitoring. As an alternative to automated weather stations, PERSIANN can extract rainfall intensity information for each storm event recorded by CLICOM stations.

The main characteristic of these data sources is that they are free to use and available for the entire Mexican territory.

3.3. Phase 3: Hydrometeorological risk forecast

In this phase, implementing a Rainfall-Runoff Model (RR-Model) is contemplated. In phases 1 and 2, rainfall records were integrated with the TPW estimated on a potential deep convective event, resulting in a precipitation scenario. In this aspect, phase 3 is sequential to the other two phases.

Various hydrological and flood models can be used to estimate the risk of river or land flooding (Teng et al., 2017). However, in both cases, two main components are required: information on the attributes of the simulation area, such as basin parameters, topographic configuration, type of cover, and soil texture, and the water input, either the hydrograph of a streamflow or hydraulic infrastructure or a precipitation scenario.

In this context, the third phase of this framework has the flexibility to be implemented according to the needs of the forecaster since, in this phase, the terrain characterization focused on a hydraulic flood model is provided by phase 2. Likewise, these models are fed in real-time under different precipitation scenarios estimated from the precipitable water content in the deep convective clouds detected by the ML models and the lightning post-processing filter.

This study presents an example of applying this framework at the Los Mochis, Sinaloa test site (Figure 1). In this case, the peak rate of discharge was estimated under a precipitation scenario generated from the simulation of the Tropical Depression 19-E event, which had significant impacts in the study area in September 2018, and using the parameters of the basin where the urban center of Los Mochis is located (Dávila Ortiz, 2019). For the estimation of the input hydrograph, the methodology of the Soil Conservation Service (SCS), Dimensionless Unit Hydrograph (DUH), was followed (Wanielista & Yousef, 1992). In Chapter 5, this case study is discussed in detail.

Although the framework has been tested in a specific study site, a detailed validation procedure of this methodology based on historical reference events is necessary. Nevertheless, as happens with meteorological sensors, Mexico lacks a network of hydrometric stations for continuous and reliable flow recording.

In order to preserve the main feature of the use of open-access products and open-source software in this framework, future work should couple this framework with an open-source numerical model for the simulation of hydrological events. The numerical model proposed is based on the two-dimensional Saint-Venant equations derived from the Navier-Stokes equations (Negrete Correa, 2016). This model is written in Fortran code and considers turbulence with the Eddy coefficient and friction based on Manning's formula. The method used to discretize the simulation domain is the finite difference method.

Chapter 4. Methods

4.1 Data

4.1.1. ABI-GOES Data

The Advanced Baseline Imager (ABI) is the primary instrument in the Geostationary Operational Environmental Satellite-R Series, providing high-resolution and real-time information about the Earth's atmosphere and weather conditions. This passive imaging radiometer has 16 spectral bands (Table 1), including two visible channels, four near-infrared channels, and ten infrared channels (Schmit et al., 2017). Models and tools use these different channels (wavelengths) to indicate various elements on the Earth's surface or in the atmosphere, such as trees, water, clouds, moisture, and smoke. In this study, 8 of the 16 GOES East (GOES 16) bands for the Continental US (CONUS; Figure 8) domain were used to generate 12 predictors based on the spectral properties of deep convective clouds. The temporal resolution of the ABI sensor is 5 minutes.

Table 1. Specifications of ABI-GOES Bands

ABI Bands	Bands Name	Center wavelength (µm)	Best Spatial Resolution (km)
1	Blue	0.47	1
2	Red	0.64	0.5
3	Veggie	0.86	1
4	Cirrus	1.37	2
5	Snow/Ice	1.61	1
6	Cloud Particle Size	2.24	2
7	Shortwave Window	3.9	2
8	Upper-Level Tropospheric Water Vapor	6.19	2
9	Mid-Level Tropospheric Water Vapor	6.93	2
10	Lower-level Water Vapor	7.37	2
11	Cloud-Top Phase	8.44	2
12	Ozone	9.61	2
13	Clean IR Longwave Window	10.33	2
14	IR Longwave Window	11.21	2
15	Dirty Longwave Window	12.29	2
16	CO2 longwave infrared	13.28	2

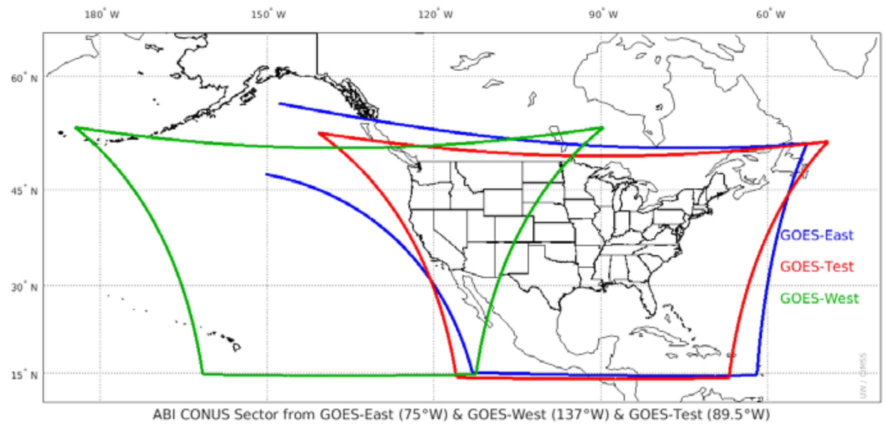


Figure 8. GOES-R CONUS domains for Operational GOES-West Location (Green), Check-out Location (Red) and Operational GOES-East Location (Blue). Source: Lindstrom (2017).

4.1.2. GLM-GOES Data

The Geostationary Lightning Mapper (GLM) and the ABI sensor form the Earth-facing or nadir-pointing payload of the GOES-R series of satellite instruments (Figure 9).

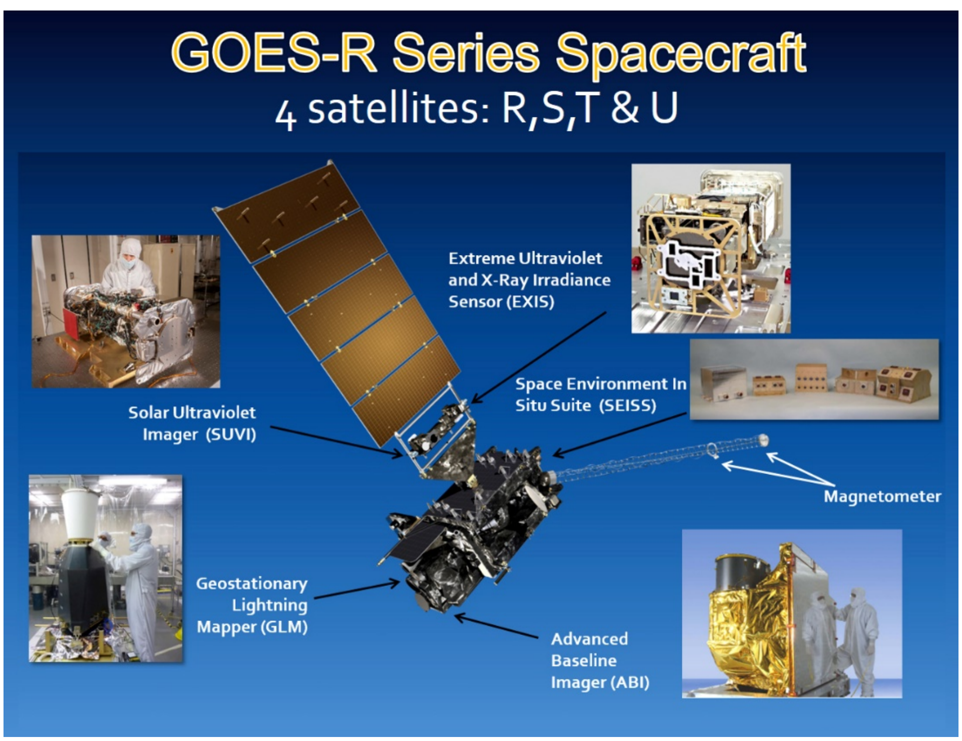


Figure 9. GOES R series instruments.
Source: https://cimss.ssec.wisc.edu/education/goesr/goesr1.html?version=mobile.

This sensor is a single-channel, near-infrared optical transient detector capable of detecting transient changes in an optical scene that indicate the presence of lightning. The GLM sensor continuously captures high-resolution images of lightning, including cloud-to-ground, intra-cloud, and cloud-to-cloud lightning, day and night, with a near-uniform spatial resolution of 8 km at a product refresh rate of less than 20 seconds over the continental United States and adjacent oceanic regions (Goodman et al., 2013).

Intense convective activity is closely associated with the incidence of lightning strikes. Hence, a post-processing filter was generated using the presence of lightning activity as an indicator of CC class. A detailed description of this post-processing filter will be provided in Section 4.2.7.

4.1.3. ABI-GOES derived products

The ABI sensor collects data across different scales, ranging from CONUS to Full Disk (providing a comprehensive view of Earth's disk from the satellite's perspective) and two mesoscale domains. This information is utilized to generate various products with a wide range of applications, including forest fire detection and analysis, volcanic and hurricane monitoring, the study of atmospheric aerosols, and detection, development, characterization, and evolution of clouds, precipitation, and lightning activity (Table 2).

Table 2. Specifications of ABI-GOES derived products.

ID	Data Product	ABI Bands (CMIP) used	ABI derived products used	Best Spatial Resolution	Temporal resolution
CMIP	Cloud and Moisture Imagery	1-16	--	See Table 1	5 min
CSM	Clear Sky Mask	2, 4, 5, 10, 11, 14, 15	--	2km	15 min
CPSD	Cloud Particle Size Distribution	2, 6	CSM, CP	2 km	5 min
COD	Cloud Optical Depth	2, 6	CSM, CP, CPSD	2 km	5 min
CP	Cloud Phase	10, 11, 14, 15	Cloud Type	2 km	5 min
CtH	Cloud-top Height	14, 15, 16	CSM	10 km	5 min
CtP	Cloud-top Pressure	14, 15, 16	CSM	10 km	5 min
CtT	Cloud-top Temperature	14, 15, 16	CSM	2 km	15 min
DSI	Derived Stability Indices	8, 9, 10, 11, 12,13, 14,15, 16	--	10 km	5 min
DMW	Derived Motion Winds	2, 7, 8, 9, 10, 14	--	10 km	15 min
TPW	Total Precipitable Water	8, 9, 10, 11, 12,13, 14,15, 16	CSM, CPSD, COD	10 km	5 min

The GOES-derived products used in this research to characterize the identified vertically developing clouds and generate precipitation scenarios are Cloud and Moisture Imagery (CMIP), Cloud-top Height (CtH), Cloud-top Temperature (CtT), and Cloud-top Pressure (CtP), Cloud Optical Depth (COD), and Total Precipitable Water (TPW).

The CMIP consists of the 16 ABI bands radiometrically calibrated and converted to brightness values (BVs) and brightness temperatures (BTs; see Table 2) in three coverage areas: Full Disk, CONUS, and two mesoscale domains in a NetCDF file (Figure 10; NOAA & NASA program, 2019).

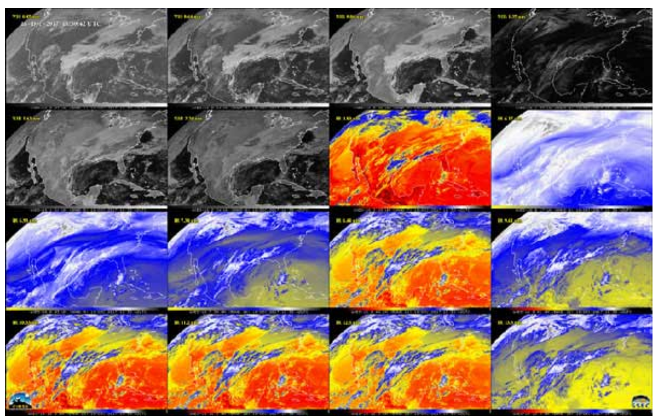


Figure 10. GOES-16 ABI imagery for each of the instrument's 16 bands (CONUS) on December 18, 2017. Source: https://www.goes-r.gov/products/baseline-cloud-moisture-imagery.html.

The Cloud Top Height algorithm uses the ABI infrared bands to retrieve the cloud-top properties simultaneously: CtH, CtT, and CtP for each cloudy pixel. This algorithm extracts the desired cloud top information from the ABI's infrared observations. Infrared observations are influenced not only by the height of the cloud but also by its emissivity and how it varies with wavelength (a behavior related to cloud microphysics). In addition, emissions from the surface and atmosphere can significantly contribute to the observed signal (NOAA & NASA program, 2019).

These cloud products are a prerequisite for generating other downstream products, including the Cloud Layer, Cloud Optical/Microphysical, and Derived Motion Wind (DMW) products. This information can be used to determine areas of cloud growth and the likelihood of precipitation.

The Cloud Type product includes six cloud classifications, whereas the Cloud Phase includes four. Cloud Type categories include warm (> 0° C) liquid water clouds, supercooled liquid water (< 0° C), mixed phase, opaque ice, cirrus (semi-transparent ice clouds), and multi-layer clouds (with a semi-transparent top layer). Cloud Phase categories include the warm liquid water phase, supercooled liquid water phase, mixed phase, and ice phase. The Cloud Phase is derived directly from the Cloud Type categories. The Cloud Type product provides valuable information on multi-layer clouds and cirrus, which benefits higher-level algorithms such as cloud top height retrieval. The Cloud Type algorithm use four ABI infrared spectral bands (NOAA & NASA program, 2019).

The COD is a dimensionless quantity that expresses the ratio of the amount of light absorbed or scattered by a cloud to the amount of incident light. The COD product uses visible and near-infrared ABI bands during the daytime and a combination of infrared bands for nighttime detection. This product, along with the Cloud Particle Size Distribution (CPSD) product, provides valuable information about the radiative properties of clouds (NOAA & NASA program, 2019).

The DMW product is derived using a sequence of visible or infrared spectral bands to track the motion of cloud features and water vapor gradients. The resulting estimates of atmospheric motion are assigned heights using the cloud height product (NOAA & NASA program, 2019).

The Derived Stability Indices (DSI) product, including the Convective Available Potential Energy, Lift Index, Total Totals, Showalter Index, and K Index, are calculated based on retrieved atmospheric moisture and temperature data. These indices help forecasters predict severe weather events by visually representing atmospheric stability parameters. Forecasters use this data to track dynamic shifts in stability at different locations, increasing their awareness of pre-convective conditions and aiding in identifying potential watch/warning scenarios (NOAA & NASA program, 2019).

The TPW contains an image with pixel values that indicate the amount of integrated column water vapor from the surface to a height corresponding to an atmospheric pressure of 300 hPa (approximately the height at the top of the atmosphere). The units of measurement for the TPW value are millimeters (mm; NOAA & NASA program, 2019).

4.1.4. MODIS Data

The Moderate-Resolution Imaging Spectroradiometer MODIS is a passive imager mounted on both the Terra (launched by NASA in 1999) and Aqua (launched in 2002) sun-synchronous polar-orbiting satellites. This sensor collects data in 36 spectral bands ranging in wavelength from 0.4 μm to 14.4 μm and at varying spatial resolutions (2 bands at 250 m, 5 bands at 500 m, and 29 bands at 1 km). Together, the instruments image the entire Earth every 1 to 2 days.

MODIS provides pixel-level retrievals of cloud top and optical properties (Platnick et al., 2017). In this research work, the variables "Cloud_Top_Preassure" and "Cloud_Optical_Thickness" (related to the ABI-GOES derived product COD) from Collection 6 (C6) of MODIS products were collected for target label generation. MODIS C6 introduces several improvements and refinements over previous versions of the MODIS data products (Dávila Ortiz et al., 2023a; Dávila-Ortiz et al., 2024).

4.1.5. PERSIANN Data

The operational PERSIANN system developed by the Center for Hydrometeorology and Remote Sensing (CHRS; Nguyen et al., 2019) uses neural network-based classification/approximation procedures. These techniques estimate the precipitation rate for each 0.25° x 0.25° pixel at an hourly temporal resolution within the infrared brightness temperature image received from geostationary satellites. The PERSIANN algorithm in this context relies on geostationary long-wave infrared imagery to generate a global rainfall dataset. The resulting precipitation product covers a global range from 60°S to 60°N (CHRS, 2023).

4.2. Methodology: Deep convective events identification system

4.2.1. Preparation of Training Dataset

Combinations of 8 ABI GOES bands were computed to create "Interest Fields" predictor variables. Deep convective cloud detection algorithms use a variety of Interest Fields tailored for different sensors such as ABI GOES (Mecikalski et al., 2015), Advanced Himawari Imager (AHI; Han et al., 2019; Lee et al., 2017), the Meteorological Imager Onboard Communication, Ocean, and Meteorological Satellite (COMS-MI; H. Han et al., 2015), and others. These predictor variables capture a range of physical properties of clouds, including CtT, cloud-top cooling rate, COD, and CtH. For this study, the Interest Fields are taken up in the works of Lee et al. (2017) and D. Han et al. (2019), listed in Table 3.

Table 3. Overview of the input data for the implementation of ML workflow and subsequent post-processing filter. The central wavelength is provided in parentheses. The initial choice of predictor variables (Interest Fields) was adopted from the work by Lee et al. (2017).

ID	Feature names	Type	Sensor
CtT	Cloud-top temperature (**11.2 μm TB**)	Predictor	ABI
CtH01	Cloud-top height 01 (**6.2 - 11.2 μm**)	Predictor	ABI
CtH02	Cloud-top height 02 (**6.2 - 7.3 μm**)	Predictor	ABI
CtH03	Cloud-top height 03 (**13.3 - 11.2 μm**)	Predictor	ABI
CtG01	Cloud-top glaciation 01 (**12.3 - 11.2 μm**)	Predictor	ABI
CtG02	Cloud-top glaciation 02 (**8.6 - 11.2 μm**)	Predictor	ABI
CtG03	Cloud-top glaciation 03 (**8.6 - 11.2 μm**) - (**11.2 - 12.3 μm**)	Predictor	ABI
CtCrate	Cloud-top cooling rate (**11.2 μm time trend**)	Predictor	ABI
TChCtH01	Temporal changes in cloud-top height 01 (**6.2 - 11.2 μm time trend**)	Predictor	ABI
TChCtH02	Temporal changes in cloud-top height 02 (**6.2 - 7.3 μm time trend**)	Predictor	ABI
TChCtH03	Temporal changes in cloud-top height 03 (**13.3 - 11.2 μm time trend**)	Predictor	ABI
TChCtG03	Temporal changes in cloud-top glaciation (**(8.6 - 11.2 μm) - (11.2 - 12.3 μm) time trend**)	Predictor	ABI
LF	Lightning filter	Filter Array	GLM
CC_labels	Deep convective cloud labels	Target Variable	MODIS

At the same time, Target Labels were created using the "Cloud_Optical_Thickness" (COT) and cloud_top_pressure_1km (CtP) variables extracted from MODIS C6. The criteria for identifying CC within the MODIS survey were based on the widely accepted classification of the International Satellite Cloud Climatology Project (ISCCP; Rossow & Schiffer, 1991). In this classification, cells with COT values greater than 23 and CtP values less than 440 mb are referred to as deep convective clouds (Figure 11).

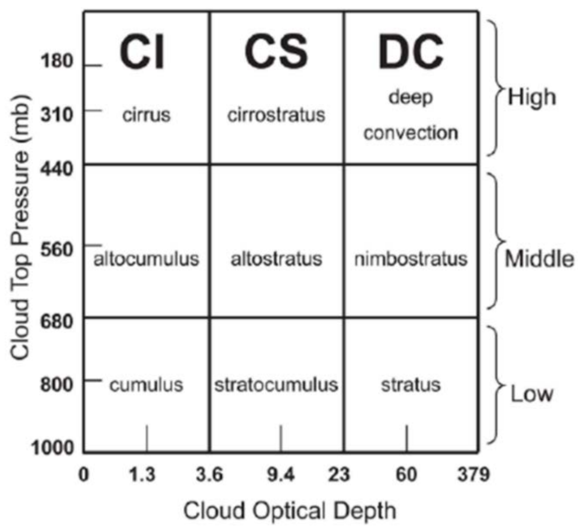


Figure 11. ISCCP Cloud Classification in terms of Cloud Top Pressure and Cloud Optical Depth. Source: Rossow & Schiffer (1991).

The principal aim of the methodology is the construction of a reference event dataset for each test site. This dataset, compiled from all available MODIS images from 2018 to 2022 between the highest rainfall months from May to September, provides a standardized and balanced data source for training and testing the different ML approaches.

MODIS images were cropped according to the specific domain of the study areas, and only those regions with a labeled convective cell (CC) percentage greater than 30% were selected. This selection of MODIS scenes was performed to create a Target Label dataset, ensuring that ML models were not trained on unbalanced class distribution data. All identified sections were then uniformly resampled using the pyresample function, a Python package designed for resampling geospatial image data. This process aimed to standardize the images with the spatial configuration of the 12 Interest Fields arrays, each consisting of 50 x 50 cells in both domains (Figures 1 and 3).

The training dataset for Mexico City was generated with 24 reference events (60,000 examples), while the test dataset had 31 events (77,500 examples). For Los Mochis, 6 (15,000 examples) and eight events (20,000 examples) were used alternately for the training and test datasets. Figure 12 shows the distribution of datasets for the Mexico City and Los Mochis study sites.

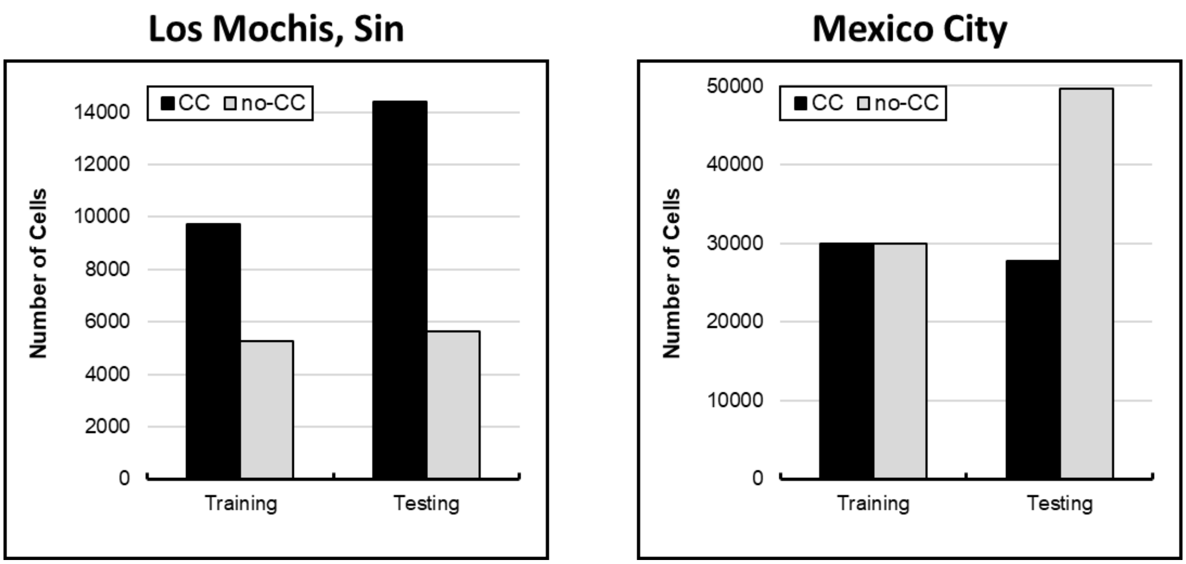


Figure 12. Distribution of Los Mochis, Sin and Mexico City datasets. For Los Mochis dataset, the number of classes labeled as CC is 24,116 and noCC is 10,884, whereas for the Mexico City dataset, the number of classes labeled as CC is 57,833 and noCC is 79,667.

4.2.2. Machine Learning Approaches

In this study, eight different ML approaches are compared (Table 4), such as LR, RF, and MLP, which are widely used in convective hazard forecasting (*e.g.*, Ahijevych et al., 2016; Burke et al., 2020; Jergensen et al., 2020; La Fata et al., 2021, 2022; Lee et al., 2017; McGovern et al., 2023; Mecikalski et al., 2015; Steinkruger et al., 2020; Yao et al., 2020) along with ML approaches based on Ensemble Learning techniques (*i.e.*, a group of predictors, called an ensemble, are trained together to improve predictive ability; Figure 13). All ML models, assembly strategies, metric estimation, and preprocessing methods were taken from scikit-learn, a free software ML library for the Python programming language (https://scikit-learn.org/stable/).

Table 4. ML models and approaches compared in this study.

Models and Ensemble Methods	Abbreviation
Logistic regression	LR
Decision tree	DT
Random forest	RF
Support vector machine	SVM
Multi-layer perceptron	MLP
Bagging (Logistic regression)	Bagging
Stacking (Logistic regression)	LRstacking
Stacking (Random Forest)	RFstacking
Hard Voting	Hvoting
Soft Voting	Svoting

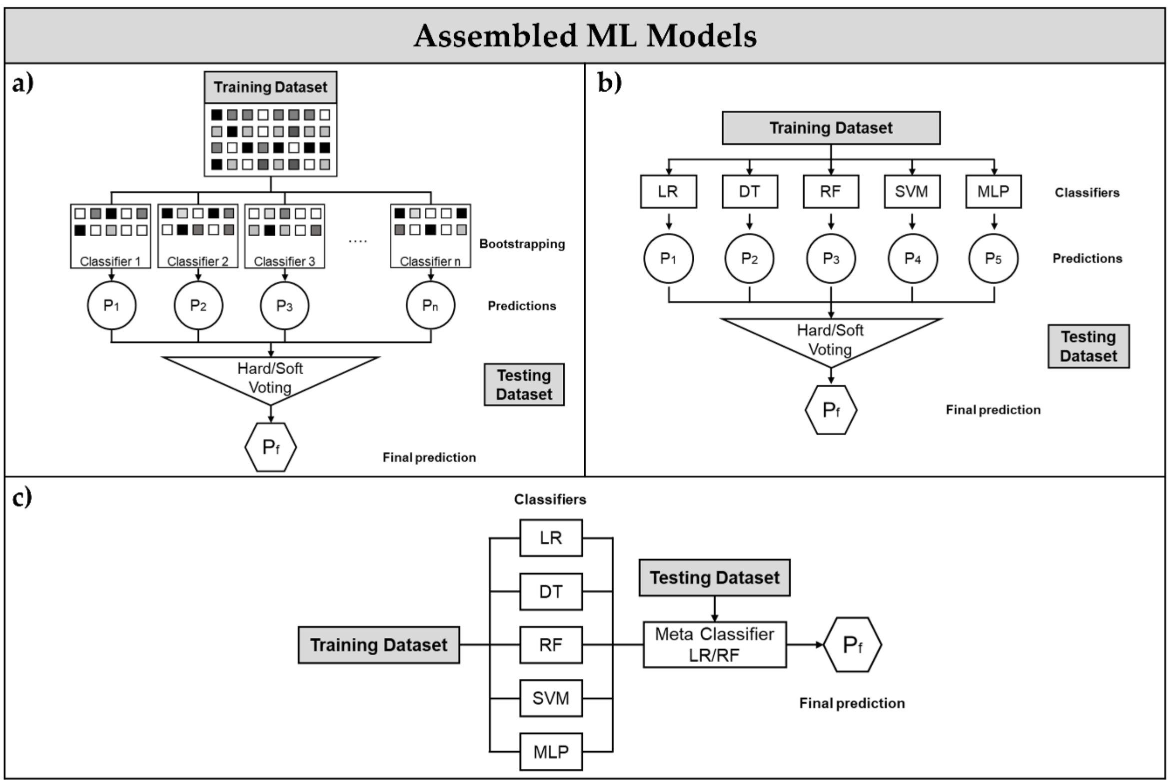


Figure 13. Diagram of ensemble ML methods. a) Bagging, b) Voting, c) Stacking.

- Logistic Regression (LR)

LR is a statistical method used to predict a binary variable (here, CC or noCC class) from one or more independent variables or predictors (Hosmer et al., 2013). LR works by fitting a logistic function to the input data, which maps the input features to the output class probabilities (Figure 14). The output of the logistic regression model is a value between 0 and 1, representing the probability that the input belongs to the positive class. This output is then the threshold for making the final binary classification decision. The LR formula is given by Equation 1.

$$E(Y) = \frac{1}{1 + exp\left[-\left(\beta_o + \sum_{j=i}^{k} \beta_j X_j\right)\right]}, \qquad (Eq. 1)$$

where, *E* is the expected value of the dependent variable *Y*, which has values of 0 and 1, equivalent to noCC and CC, respectively, *k* is the number of predictors, and *Xj* is the value of the *j* th predictor. The parameters *βo,…, βj* are linear coefficients or weights for the predictor variables (Dávila-Ortiz et al., 2024).

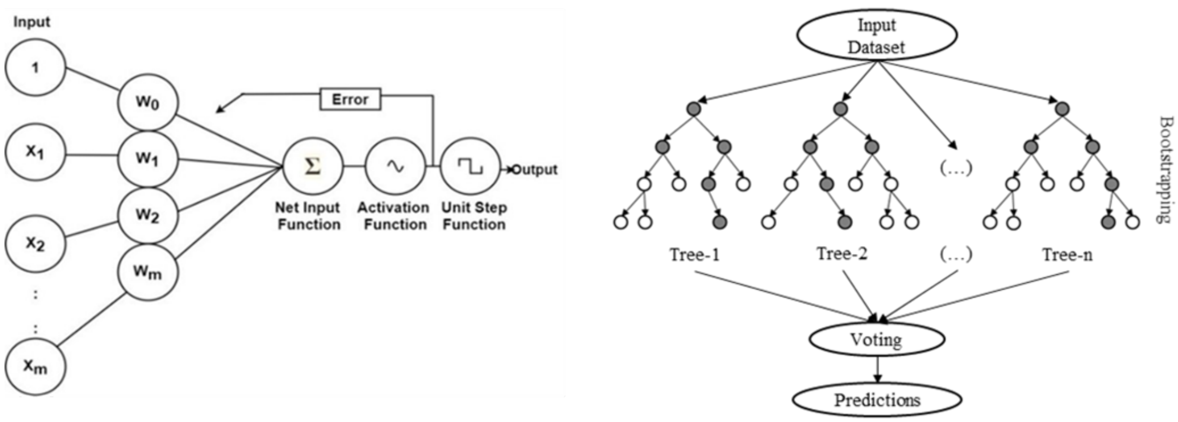


Figure 14. Architecture of a Logistic Regression (left; Torres et al., 2019) and Random Forest (right; Sahour et al., 2021).

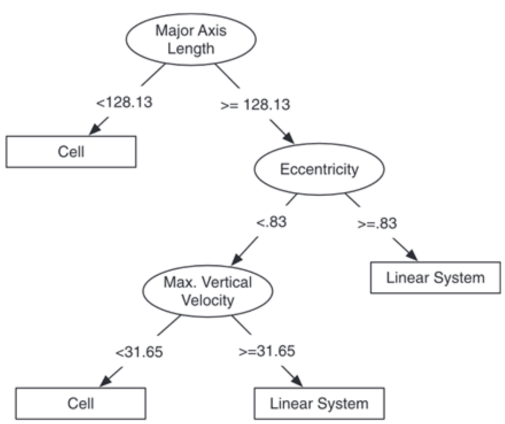
The logistic regression model undergoes training by optimizing a cost function, typically the Cross-Entropy Loss. This is achieved through techniques like Gradient Descent, Stochastic Average Gradient, Newton method, etc. The model's parameters are iteratively adjusted to minimize the disparity between the predicted probabilities and the actual class labels in the training data (Hosmer et al., 2013).

A key advantage of logistic regression is its simplicity and interpretability. The model provides insight into the contribution of each feature to the prediction, making it easier to understand and communicate the classification results. Despite its simplicity, LR performs well on several real-world problems, including non-linear problems in atmospheric science such as convection (Ukkonen & Mäkelä, 2019).

- Decision Tree (DT)

DT is a nonparametric supervised learning method that recursively partitions the dataset into subsets based on the values of the input features, ultimately assigning a label to each observation (Breiman, 1984). DTs are constructed in a hierarchical structure, where each node represents a decision based on a particular feature, and each branch represents the possible outcome of that decision.

Decision trees, with their interpretability and visualizability, are instrumental in understanding the decision-making process (Figure 15). However, they can be prone to overfitting, particularly when the tree is deep. To mitigate this, ensemble methods like RF are often employed, which amalgamate multiple decision trees to enhance overall performance and generalization. Notable examples of convective hazard forecasting using DT are the works of Gagne et al., (2009) and H. Han et al. (2015).



Morphological	Reflectivity	Wind	Control
Eccentricity	Max	Max speed	Mean stratiform distance
Major axis length	Min	Min speed	Storm direction
Minor axis length	Mean	Mean speed	
Orientation	Std dev	Max vertical velocity	
Equivalent diameter	Range	Min vertical velocity	
Area		Mean vertical velocity	

Figure 15. An example of the general-type decision tree used by Gagne et al. (2009) to classify convective areas using DTs (left), and the attribute sets used (right).

- Support Vector Machine (SVM)

Developed by Vapnik (1963), SVM is a robust supervised machine learning algorithm that can be used for both classification and regression tasks. SVM works by finding the optimal hyperplane in a high-dimensional space to separate different classes. The key objective is maximizing the margin, which is the distance between the hyperplane and the nearest data points from each class, called support vectors

(Figure 16). The SVM aims to balance maximizing this margin and minimizing misclassifications. The algorithm can handle non-linear decision boundaries using kernel functions that transform the input features into a higher-dimensional space. The SVM classifier is trained to find the hyperplane that best separates the data, making it a robust and effective tool for binary and multi-class classification tasks. SVM is particularly effective in classifying complex small- or medium-sized datasets and high-dimensional spaces (Géron, 2019).

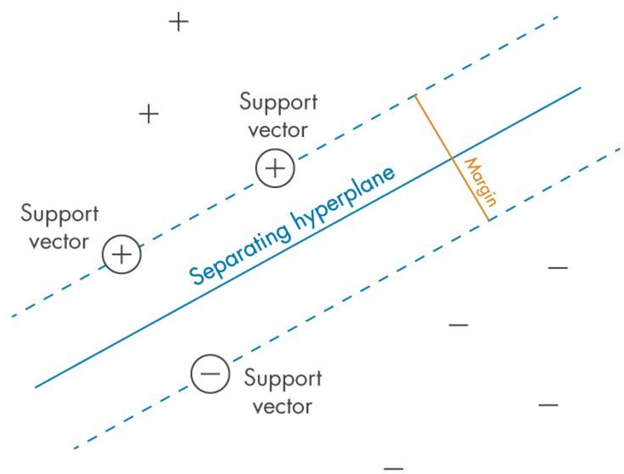


Figure 16. Definition of the "margin" between classes: the criterion that SVMs try to optimize. Source: https://la.mathworks.com/discovery/support-vector-machine.html.

SVM is widely used in various convective hazard forecasting applications, for example, in tornado prediction (Adrianto et al., 2009; Trafalis et al., 2003) or convective thunderstorm forecasting (L. Han et al., 2017; Jergensen et al., 2020; Sangiorgio et al., 2020).

- Multi-layer Perceptron (MLP)

The MLP is a variant of an artificial feedforward neural network (ANN; the first conceptual model of an artificial neural network was developed by McCulloch & Pitts (1943). The MLP consists of three layers of interconnected nodes called neurons. These layers have specific functions that allow the model to learn complex non-linear relationships between input and output variables. The input layer receives raw input data, with the number of neurons determined by the number of features or predictors.

The hidden layer is responsible for extracting relevant features and learning complex patterns. Finally, the output layer classifies observations based on their class probability. In this case, the output layer consists of two neurons, reflecting a binary classification task that distinguishes between CC and noCC classes (Figure 17).

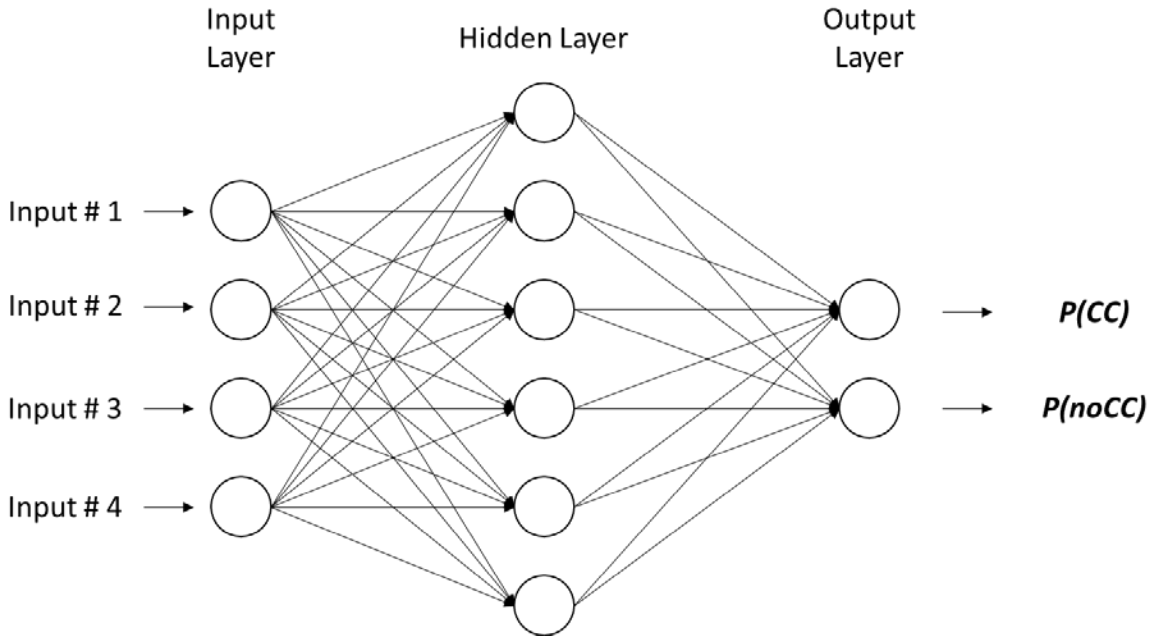


Figure 17. Example of MLP Network adapted for the CC and noCC classification task.

MLPs have been widely used in numerous studies on convective hazard and precipitation forecasting (*e.g.*, Chen et al., 2020; Sobash et al., 2020). Nevertheless, there has been a predominant shift toward the use of advanced deep neural network architectures in this field, including Deep Neural Networks (DNN; *e.g.*, Afzali Gorooh et al., 2020; Y. Lee et al., 2019; Liu et al., 2019), Recurrent Neural Networks (RNN; *e.g.*, Akbari Asanjan et al., 2018; Leinonen et al., 2022), and CNN (*e.g.*, Hilburn et al., 2020; Kim et al., 2018; Y. Lee et al., 2021).

- Bagging and Random Forest (RF)

Bootstrap Aggregation, or Bagging, is a consensus-based ML technique used to improve accuracy, reduce variance, and avoid overfitting, resulting in a more robust and reliable predictive model. By combining multiple models trained on different subsets of the training data derived by bootstrap sampling, where data points are randomly selected with replacement (Figure 13a; Breiman, 1996). A variant of this method, in which sampling is performed without replacement, is called pasting (Breiman, 1999; Géron, 2019).

Introduced by Breiman (2001), RF is a nonparametric ensemble model that takes the Bagging approach by constructing many decision trees during the training phase and combining their predictions during the testing phase (Figure 13a). The outputs of the individual trees are then aggregated by majority voting to produce the final prediction. The RF measures feature importance, indicating which features contribute more significantly to the model's predictive performance. This information is valuable for feature selection in data analysis. In addition, training individual trees in an RF can be performed in parallel, making it computationally efficient and suitable for large datasets.

The random forest (RF) model has found extensive practical applications in the field of convective hazard forecasting, making it among the most widely used machine learning models, along with LR. Their popularity stems from their ability to capture complex non-linear relationships between predictors and predictands, such as convective storm systems or precipitation (Ramirez & Lizarazo, 2017). Notable examples of their applications in the literature include the works of Ahijevych et al. (2016), Burke et al. (2020), H. Han et al. (2015), Jergensen et al. (2020), Kim et al. (2017), La Fata et al. (2021), S. Lee et al. (2017), Liu et al. (2019), Mecikalski et al., (2015), Steinkruger et al. (2020) and Yao et al. (2020).

In this study, two Bagging approaches were implemented, RF and Bagging method used LR as predictor in the bootstrapping samples instead of DTs (henceforth referred to as Bagging; Table 4 and Figure 13a).

- Voting

Voting is an assembly strategy used to make predictions by combining the collective outputs of multiple independently trained models or classifiers in parallel on the same dataset. In this study, the ML approach involves aggregating the predictions of each classifier (LR, DT, RF, SVM, and MLP; Figure 13b) and determining the class with the highest number of votes. If the class with the most votes is selected as the final prediction, it is referred to as the hard voting classifier (Hvoting). Alternatively, when the class with the highest average class probability is chosen, it is termed a soft voting classifier (Svoting; Géron, 2019).

Voting mitigates the effects of individual model biases and errors, thereby improving overall accuracy and robustness. For example, in S. Lee et al. (2017) work, majority voting effectively removes salt-and-pepper noise in his results for convective initiation objects predicted with DT and RF.

- Stacking

Stacking, also known as stacked generalization or stacked ensemble, is an advanced ensemble learning technique that significantly improves predictive performance by combining the strengths of multiple models (Wolpert, 1992). In this assembly strategy, the aggregate models, each trained independently and called base models (Table 4), are used according to their weights to produce an output that a Meta-Classifier takes as input (Figure 13c). The essential advantage of this method is that it allows the Meta-Classifier to learn to optimally combine the predictions of the base models, thereby leveraging their individual strengths.

Each prediction from the base models becomes a new feature in the Meta-Classifier Dataset. The primary purpose of the stacking method is that the Meta-Classifier learns to optimally combine the predictions of the base models, taking advantage of their strengths. This work used LR and RF as Meta-Classifiers because both models reported the best performance metrics (LRstacking and RFstacking). Although not considered in this study, proper tuning of hyperparameters for both base models and the meta-model is essential to achieve optimal performance.

4.2.3. *Hyperparameters Tuning*

In ML, internal model parameters are learned directly from the data during training. However, models also have various hyperparameters that control complexity and regularization. These parameters need to be specified in advance and can strongly impact performance (Ukkonen & Mäkelä, 2019).

There are various methods for hyperparameter tuning, including manual tuning, grid search, random search, and more advanced techniques (Figure 18). Here, hyperparameter tuning was performed using GridSearchCV, a function provided by the scikit-learn library (*i.e.*, a free software ML library for the Python programming language https://scikit-learn.org/stable/) which performs an exhaustive search over a specified parameter grid to find the optimal combination of hyperparameters for an ML model (Table 5). In addition, a cross-validation strategy was used during the hyperparameter search, which was selected 5 k-fold (Figure 19).

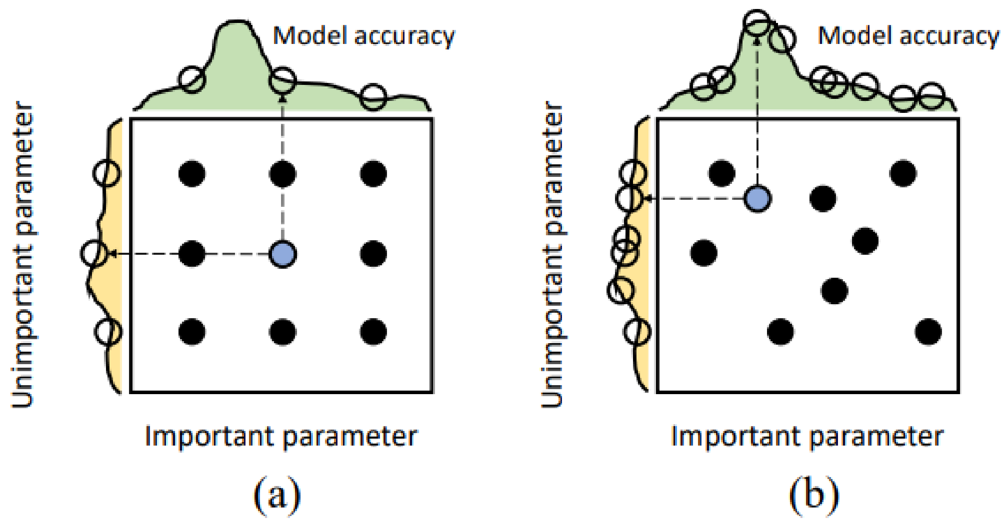


Figure 18. Comparison between (a) grid search and (b) random search for hyperparameter tuning. The nine dots represent candidates. The curves on the left and top denote the model accuracy (*e.g.*, NMSE or overall accuracy) as a function of each search dimension. Source: Salgado Pilario et al.(2021)

Table 5. Hyperparameters of the LR, RF, and MLP tuned using the cross-validation strategy.

Hyperparameter	Explored space
Norm of the penalty (LR)	L1, L2
The inverse of regularization strength (LR)	0.01, 0.1, 1, 10, 100
Solver (LR)	liblinear, lbfgs, newton-cg
Maximum depth of the tree (RF)	None, 2, 5, 10, 20, 30
Number of trees in the forest (RF)	100, 500, 1000
Random split predictor variables (RF)	1, 2, 3, 4, 5
Number of neurons in the hidden layer (MLP)	10, 100, 1000
Solver (MLP)	lbfgs, Adam, sgd

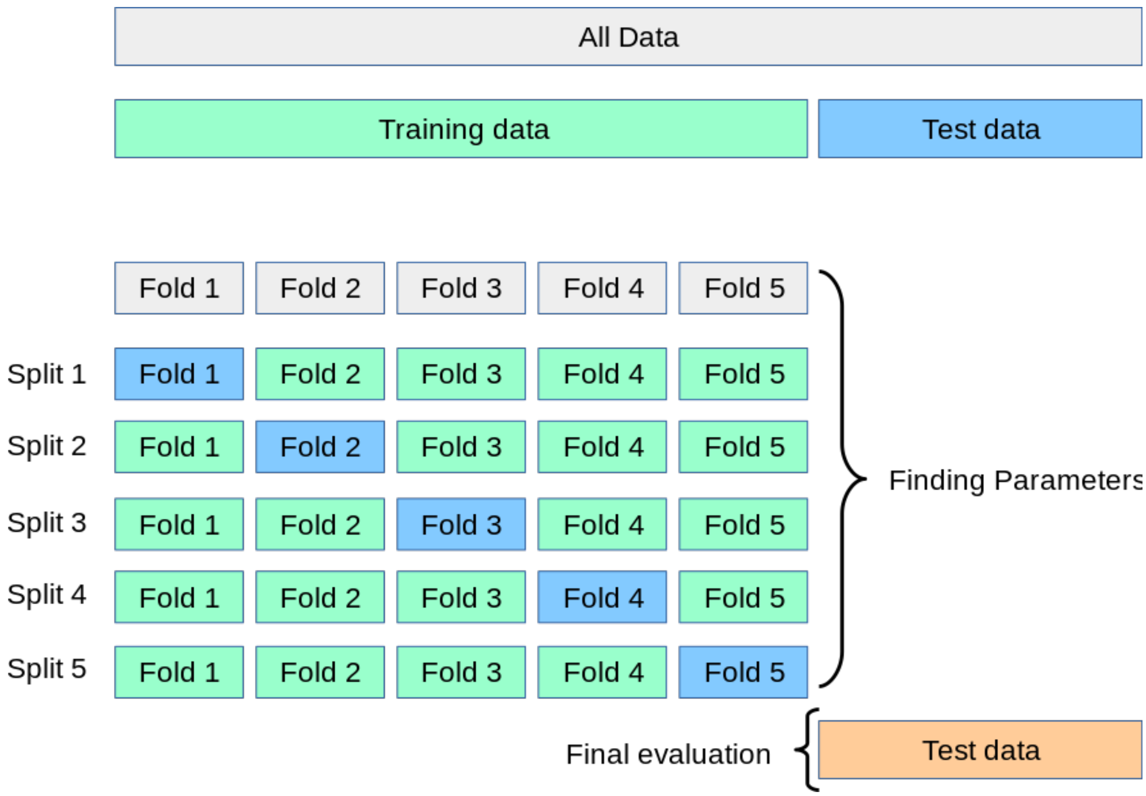


Figure 19. In cross-validation, the dataset is divided into k equally sized smaller subsets, called folds. A model is trained using k-1 of the folds as training data, leaving one for testing in each iteration, providing a robust evaluation of a machine learning model's performance. Source: https://scikit-learn.org/stable/modules/cross_validation.html.

4.2.4. Feature Scaling

Feature scaling, a method used to normalize the range of independent variables or features of data (here, interest fields), is an important preprocessing step in machine learning. The StandardScaler method in the scikit-learn library stands out for its simplicity and effectiveness. It provides a straightforward way to standardize features individually by transforming them to have $\mu = 0$ and $\sigma = 1$. This ensures uniformity and prevents any particular feature from dominating the learning process (Figure 20). This method's simplicity and effectiveness make it a popular choice for normalization in many machine-learning algorithms.

Standardization is performed by calculating the z-score for each observation in a column. Z-score, μ, and σ formulas are given by Equations 2, 3, and 4.

$$Z = \frac{\mathcal{X} - \mu}{\sigma} \qquad (Eq. 2)$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(\mathcal{X}_i) \qquad (Eq. 3)$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathcal{X}_i - \mu)^2} \qquad (Eq. 4)$$

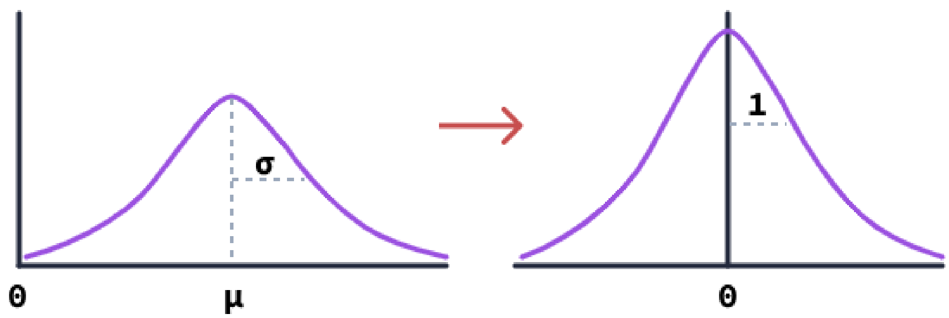


Figure 20. Schematization of the StandardScaler method. Standardize features by removing the mean and scaling to unit variance. Source: https://app.dataquest.io/m/744/k-means-with-scikit-learn-and-interpreting-results/3/scaling-the-data.

4.2.5. Machine Learning Process

To compare the performance of the eight ML approaches, uniform dataset configurations were used for training and testing in each study area. Furthermore, a feature or relative variable importance analysis was performed to estimate the degree of contribution of each Interest Field in the prediction stage.

RF and LR provide the relative importance of input variables through measures such as the Mean Decrease Impurity (MDI) and the absolute value of weighting coefficients, respectively (S. Lee et al., 2017). In this case, the feature importance of the RF is provided by the "feature_importances_" function (a method of the scikit-learn library). This approach evaluates the importance of each feature by measuring how much it contributes to the reduction of impurities, in this case, the Gini impurity, as part of the iterative process of feature-based splitting.

In the context of ML, the Gini impurity is commonly used as a criterion for splitting nodes when constructing a decision tree. The Gini impurity for each node in the binary classification task was calculated using Equation 5.

$$Gini(D) = 1 - \sum_{i=1}^{c} {P_i}^2 \qquad (Eq. 5)$$

where *D* is the node, c is the number of classes, and *Pi* is the proportion of instances of class *i* in the node. The Gini impurity is minimized when all instances in the node belong to the same class (pure node) and maximized when the instances are evenly distributed across all classes (impure node).

This relative feature importance analysis helps reduce dimensionality, improve model efficiency, and mitigate the risk of overfitting by focusing on the predictor variables with the highest contribution to CC class identification. A lower-dimensionality dataset improves model performance by removing noise from lower-weight features while significantly reducing computational cost.

Los Mochis and Mexico City datasets were reduced to six predictor variables that were selected because of their significant relative importance in LR and RF. After predictor selection and data scaling, hyperparameter tuning was performed using GridSearchCV, culminating in evaluating model performance using various accuracy metrics.

4.2.6. Workflow and implementation

A quasi-real-time system for identifying and monitoring potential convective events was designed using open-source software and open-access products, automation, and scalability principles (Figure 21). This system forms Phase 1: Characterization of storm clouds of the framework for real-time flood risk assessment discussed in the previous chapter.

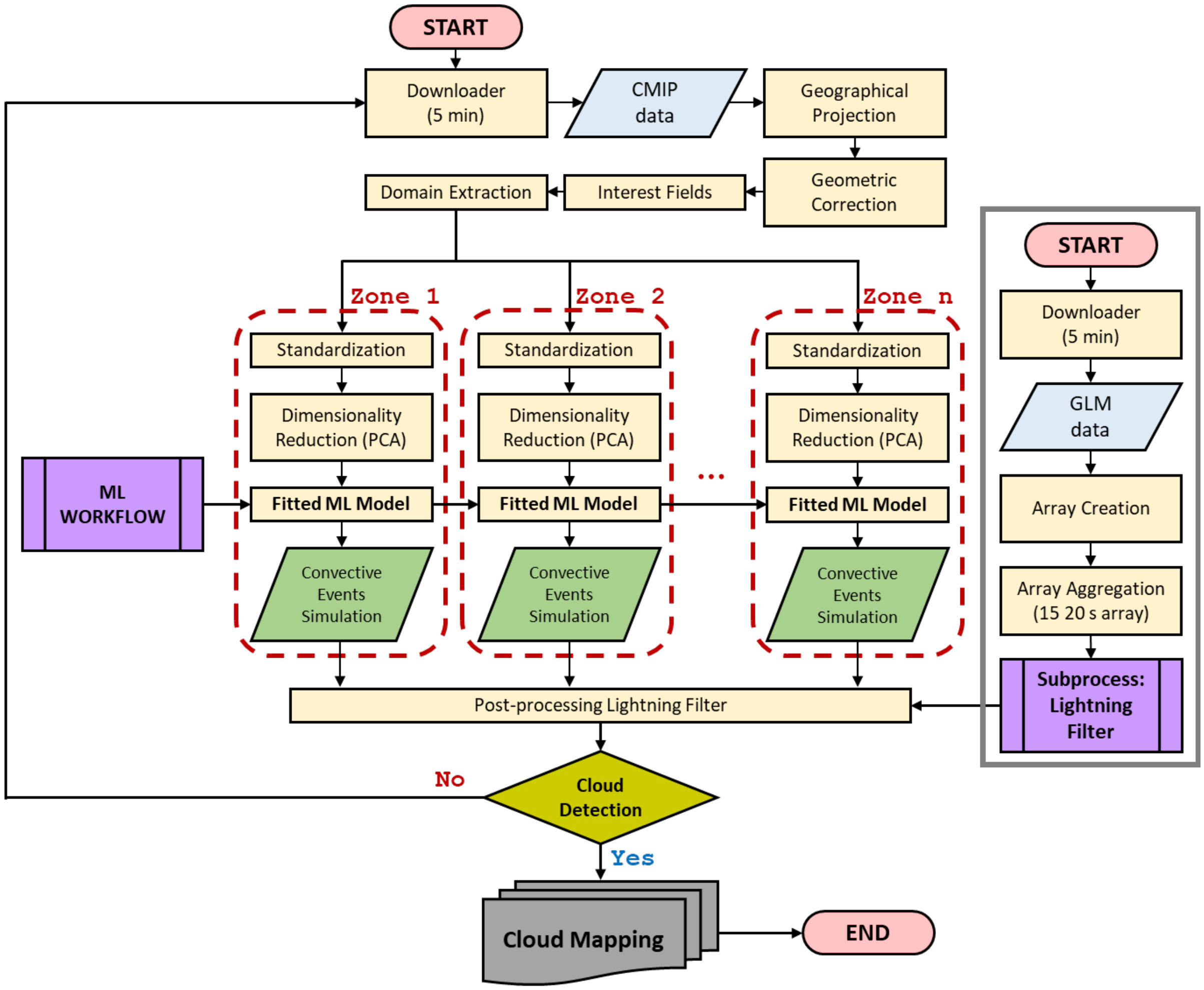


Figure 21. General scheme of the framework for the identification and monitoring of potential convective events in quasi-real time (5 min). The ML approaches used in the ML process were selected based on their performance parameters during the training and testing phases at each study site.

In accordance with the principle of using only open-access products, all processes and the workflow were implemented using Python 3 programming and designed with ABI-GOES and GLM-GOES products as inputs.

The workflow is cyclical and automated, operating every five minutes to coincide with the acquisition time of the ABI bands. It commences with the download of ABI and GLM information from Amazon Web Service (AWS) S3 storage (https://registry.opendata.aws/noaa-goes/) using Boto3, the AWS Software Development Kit (SDK) for Python. This novel methodology is designed for scalability, with the potential for spatial extrapolation to the entire Mexican country.

It follows the principles of open-access and automation, leveraging the availability of modeling inputs throughout the territory. For the training and testing phases of ML models, it is crucial to delineate the zones of homogeneous convective behavior.

The Interest Fields are calculated from the information extracted from the CMIP product, and the spatial extraction of the study areas is performed. In the second step, the pipelines (*i.e.,* a collection of sequential data manipulation operations; this approach was taken from the scikit-learn Python library) of preprocessing and modeling are fed with the Interest Fields. Previously, the models were trained and tested at each site of interest based on the creation of a set of reference historical events and a comprehensive analysis of the performance of each ML approach performance (Table 4).

Once the optimal models and hyperparameters have been identified, as well as the significant variables to be used as predictors, these configurations are stored in a pipeline (dashed red outline in Figure 21). The processing pipeline considers the scaling phase of the predictors from the StandardScaler method and the dimensionality reduction with a Principal Component Analysis (PCA).

PCA is a mathematical procedure that transforms high-dimensional data into a lower-dimensional representation while preserving its essential variance to improve computational efficiency, reduce overfitting, and result in a simpler and more interpretable model. Hasan & Abdulazeez (2021) provided a comprehensive review of the use of PCA for dimensionality reduction in a ML context.

To visualize the results, the Python Matplotlib Basemap Toolkit library is used to map the geospatial information. In addition, a post-processing filter based on lightning incidence was integrated to identify potential deep convection zones not detected by ML models. The process runs iteratively every 5 min.

4.2.7. Post-processing Lighting Filter

To strengthen the performance of the models, a post-processing filter based on lightning incidence was integrated into the CC identification and monitoring framework. According to H. Han et al. (2015), lightning is a reliable indicator of intense convective activity typical of storm cloud formation environments, making it a key component for monitoring and predicting severe weather events. Several studies related to deep convective cloud detection have included lightning data (Hilburn et al., 2020; La Fata et al., 2021, 2022; S. Lee et al., 2017; Rutledge et al., 2020; Ukkonen & Mäkelä, 2019; Zhou et al., 2020).

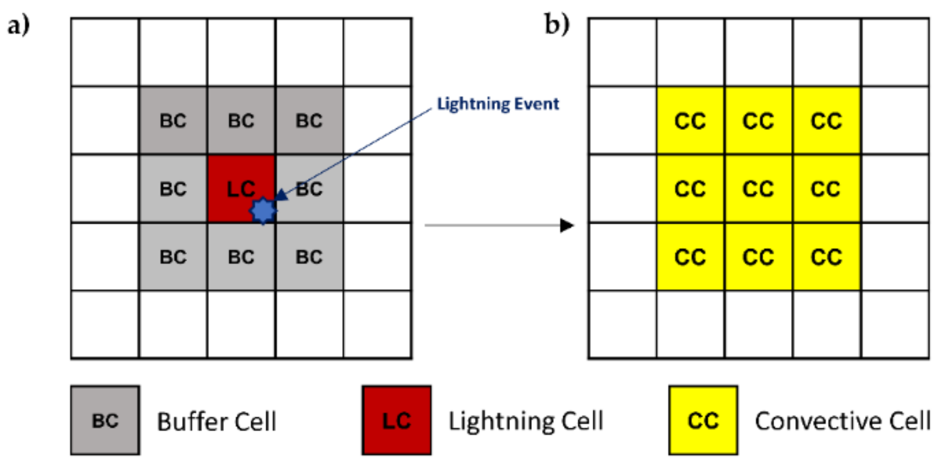


Figure 22. Schematization of the Lightning Filter generation process. When the presence of a Lightning Event is identified from the GOES GLM data accumulated in 5 min within the site of interest, it is associated with a cell of the model spatial domain, using its distance from the nearest cell center as the criterion for assignment. The Lightning Cells (LC), together with a buffer of one cell per edge (BC), are called Convection Cells (CC) and are integrated into the output matrix of the ML models.

The GLM-GOES data are used to generate a lightning matrix that is additively integrated into the ML models' output to increase the models' probability of detection (POD). The GLM variable "Event" was chosen to represent the occurrence of a single pixel exceeding the brightness detection threshold during a ~2 ms frame (Rudlosky et al., 2019). These products contain geospatial information on the occurrence of lightning with a temporal resolution of 20s. In this framework, the density or number of lightning strikes is not considered, and only the presence or absence of a meteor is used as an indicator of CC. From the spatial location where lightning is detected, an algorithm based on Euclidean distance (Equation 6) is applied to associate it with the nearest Interest Field spatial domain cell center.

$$Distance_{(LI,Ce)} = \sqrt{(Longitude_{Ce} - Longitude_{LI})^2 + (Latitude_{Ce} - Latitude_{LI})^2} \quad (Eq.6)$$

where, C_e is the cell center of the Interest Field spatial domain and *LI* is the coordinate where the lightning strike was registered. In addition, a 3x3 cell buffer was designated around the CC cell identified from the lightning filter (Figure 22).

4.2.8. Accuracy metrics

To assess the performance of the different ML approaches, well-known classification metrics were calculated from the confusion matrix, including Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), total classification Accuracy (ACC), bias (BIAS), Precision, F1 Score (F1), and Intersection over Union (IoU):

$$POD = \frac{TP}{TP + FN} \quad (Eq.7)$$

$$FAR = \frac{FP}{FP + TP} \quad (Eq.8)$$

$$CSI = \frac{TP}{TP + FP + FN} \quad (Eq.9)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (Eq.10)$$

$$Bias = \frac{TP + FP}{TP + FN} \quad (Eq.11)$$

$$Precision = \frac{TP}{TP + FP} \quad (Eq.12)$$

$$F1 = 2 \cdot \frac{Precision \cdot POD}{Precision + POD} \quad (Eq.13)$$

$$IoU = \frac{TP}{TP + FP + FN} \qquad (Eq. 14)$$

where, TP is the number of CC pixels that were correctly classified as CC (*i.e.,* true positives), FP indicates the number of CC pixels that were incorrectly detected as CC (*i.e.,* false positives), FN is the number of CC pixels that were incorrectly marked as noCC (*i.e.,* false negatives). TN is all the remaining pixels that were correctly classified as noCC. The values of TP, TN, FP, and FN were extracted from the confusion_matrix scikit learn function to generate a confusion matrix for evaluating the performance of the classification model (Figure 23). The confusion_matrix function takes the actual and predicted labels as input and returns a 2D array representing the confusion matrix.

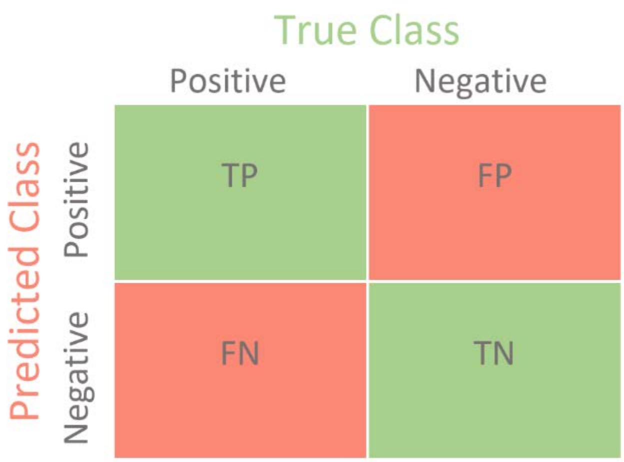


Figure 23. A confusion matrix is a table used in machine learning to evaluate the performance of a classification algorithm. It provides a summary of the model's predictions compared with the actual results, grouping the results into four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Source: https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826.

ACC ranges from 0 to 1, with 1 indicating perfect prediction and 0 indicating complete incorrectness. POD and Precision share this range, with higher values indicating better performance. FAR ranges from 0 to 1, where 0 represents no false positives and 1 indicates all positives are false. CSI ranges from -1 to 1, where 1 represents perfect prediction, 0 indicates no prediction skill, and -1 represents perfect inverse prediction. IoU ranges from 0 to 1, with higher values indicating better overlap between predicted and true areas. F1 Score also ranges from 0 to 1, with 1 representing perfect precision and recall balance. In summary, higher values are desirable for all metrics except for FAR, where lower values.

4.3. Methodology: Fusion of multi-source information

4.3.1. Dimensionless precipitation curves

Hyetographs (*i.e.,* a graphical representation depicting the variation of rainfall intensity over a specific period) and storm depth distributions are essential elements in hydraulic design, water works projects, and flood risk forecasting. Design hyetographs are used with unit hydrographs to obtain the peak discharge and hydrograph (*i.e.,* a graphical representation illustrating the variation in river discharge or streamflow over time) shape for hydraulic design (Elfeki et al., 2014). Hyetograph basics and the relationship between hyetographs and hydraulic design are discussed in Haan (1994) and Chow et al. (1994).

When it comes to the time distribution of the design storm (Intensity-Duration-Frequency), it is worth noting that the United States Soil Conservation Service (SCS) has played a significant role. They have developed four types of cumulative curves: Type I, Type II, Type III, and Type IA, which are valid for the US territory (El- Sayed, 2018). It's also common to describe these hydrographs in dimensionless forms with units of time (storm duration) and cumulative rainfall depth (storm depth) expressed as percentages of the respective totals (Elfeki et al., 2014).

This type of curves are commonly used in several studies that employ precipitation scenarios; however, Huff (1967) proposed a different approach by constructing dimensionless precipitation curves using information collected at a specific site.

'Huff curves' is a method for characterizing storm mass curves. They are a probabilistic representation of the accumulated storm depths for the corresponding accumulated storm durations, expressed in dimensionless form (Bonta, 2004). From the Huff curves, we can disaggregate precipitation amounts by fitting precipitation values to precipitation distributions generated from rainfall intensity records. The process of developing Huff curves from measured data is described in detail below. A more detailed description of this methodology can be consulted in Bonta (2004).

- Storm Identification: In order to construct Huff curves, storms must be identified and separated within a precipitation record. Although there are several approaches to identifying storms, one that is commonly used is to estimate a parameter called "Minimum Dry Period Duration" (MDPD) and then identify storm events within precipitation records (Figure 24a).

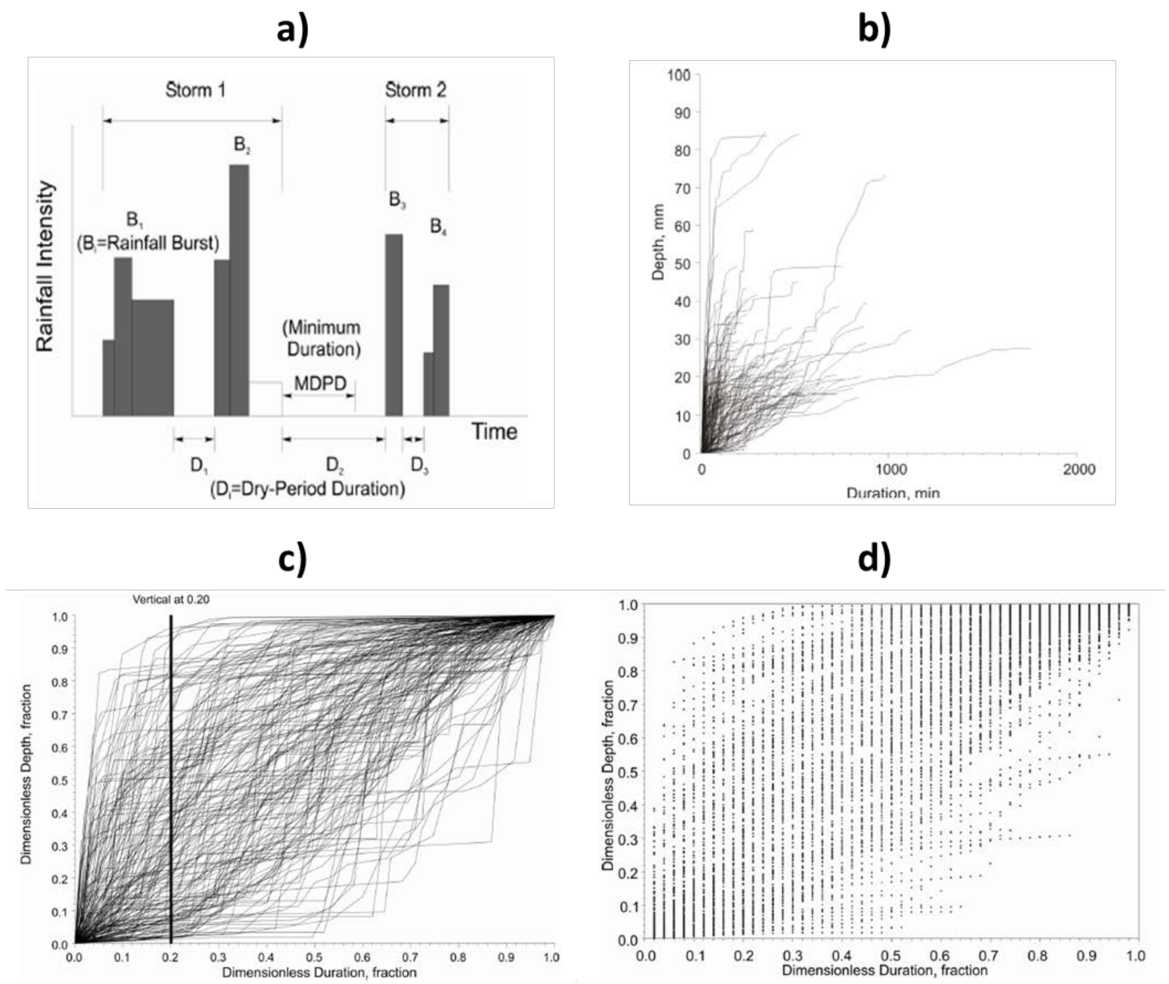


Figure 24. Schematic of the approach used to identify and separate storms (a). Storm mass curves for the identified storms (b). Dimensionless storm mass curves (c). Intersections of storm mass curves with vertical lines (d). Source: Bonta (2004).

- Storm Database: Storms are extracted from the precipitation record using monthly MDPD values to form a storm mass curve database (Figure 24b).
- Nondimensionalized Storm Curves: Each mass curve is nondimensionalized by dividing all breakpoint mass curve depths by the total storm depth and all elapsed times by the storm duration. The resulting curves are then superimposed (Figure 24c).
- Storm Curves Intersections: For a given vertical line representing a single dimensionless duration (*e.g.,* dimensionless duration−axis value of 0.2), the intersection of each dimensionless mass curve with this vertical is interpolated (Figure 24d).

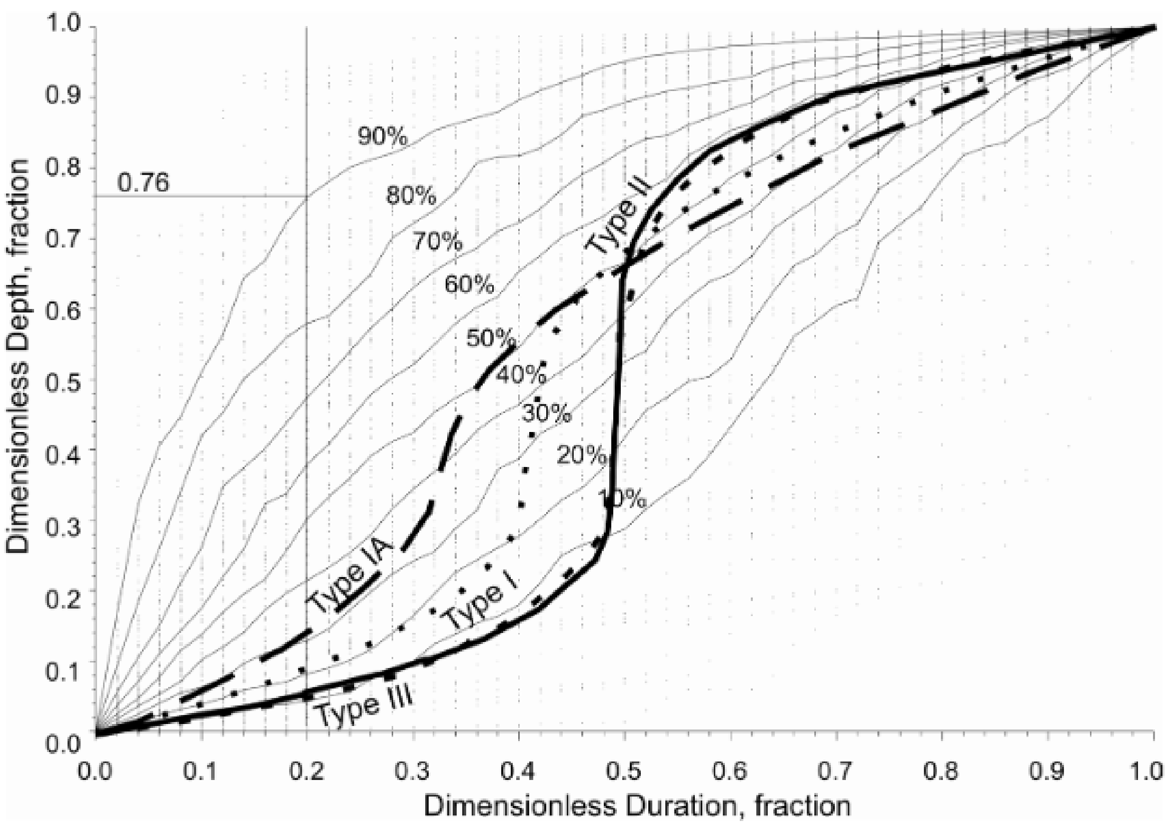


Figure 25. Storm mass−curve intersections with isopleths of probability (Huff curves) and NRCS design storms (“Type” curves) superimposed. Source: Bonta (2004).

- Building Probability Isopleths: The final step is to add the probability dimension to the graph and determine the percentage of mass curve intersections at or below the assigned percentages for each vertical. Equation 15 calculates the following percentages:

$$P = \frac{100i}{(n+1)} \qquad (Eq. 15)$$

where, *P* is the cumulative percentage of dimensionless depth points, *i* the point number and *n* the total number of points. Interpolated dimensionless depths corresponding to identical assigned probabilities are connected by straight lines, which are now called isopleths (Figure 25).

The probability isopleths comprise a set of “Huff curves”, which are a probabilistic summary representation of storm mass curves in terms of dimensionless elapsed times into a storm and corresponding dimensionless accumulated depths.

4.3.2. Extraction of deep convective cloud properties

A mask was created to extract various optical and microphysical properties of the storm clouds from the CC pixels detected with the best-fitting ML model and the addition of the post-processing lighting filter.

The array stack of each of the ABI-GOES products was spatially interpolated to have the same spatial domain as the Interest Fields generated with the CMIP data, which is approximately 2 km for the CONUS domain (Figure 26). In this instance, CtP, CtH, and TPW are resampled and interpolated because these products have a spatial resolution of ~10 km, whereas COD and CtT share the same spatial domain as the Interest Fields. In the case of the temperature variable, ABI-GOES channel 14 (IR longwave window 11.21 µm; Table 1) is used instead of CtT since it is equivalent to this GOES product, which is only available in the Full Disk domain and has different spatial and temporal resolutions than the CONUS domain products.

The pixels of the CONUS domain imagery do not have a uniform spatial resolution because cells tend to have heterogeneous spatial sizes due to the effect of data reprojection (*i.e.*, mathematical transformations that adjust the latitude and longitude geographic coordinates of each pixel or feature in the dataset to fit the desired coordinate system).

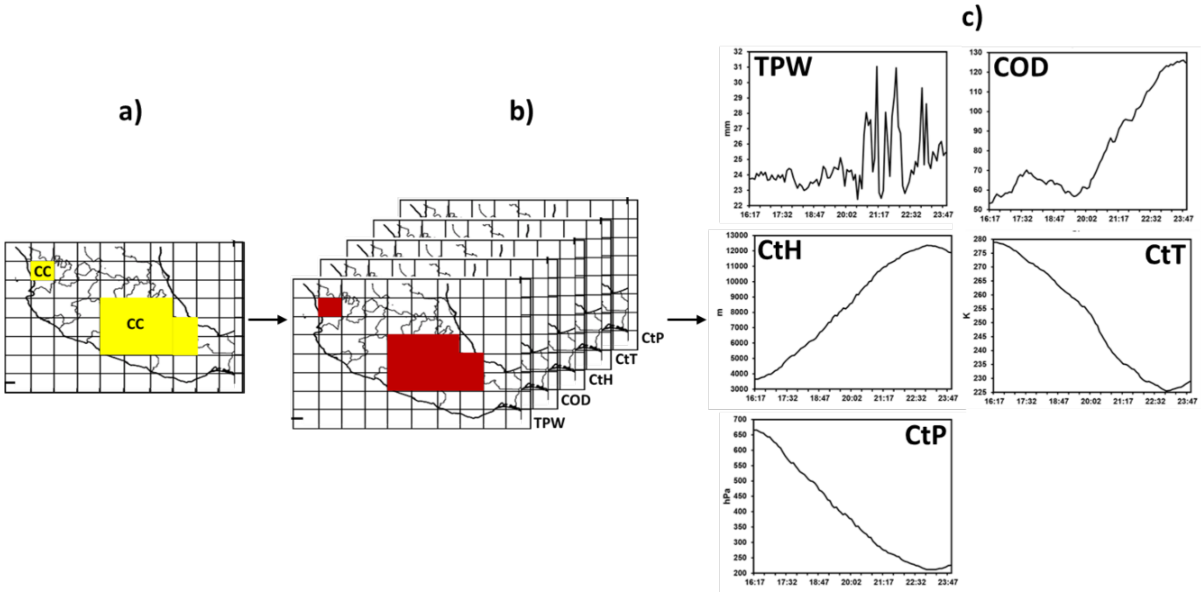


Figure 26. Schematic of the deep convective cloud property extraction process. The CC pixels are detected (a) and extraction masks are generated from them in the array stack of the GOES R products; TPW, CtH, CtP, CtT, and COD (b). The values were averaged, and time series were constructed to represent the behavior of these variables during the evolution of the potential deep convection event detected.

Once the arrays of the ABI-GOES product set are masked, the values corresponding to the zones of the study area where CC was detected are extracted. These values, which represent each variable of the simulated event(s) at a given time step, are then averaged. This average value serves as a magnitude, which we use to generate plots. These plots depict the temporal behavior of each variable in the presence of one or more potential deep convection events, providing a visual representation of the findings. In the case of TPW, this product can be used to estimate the mean amount of water that can be converted to rain in a simulated storm cloud.

4.3.3. Precipitation scenarios generation

Precipitation scenarios were generated by fitting time-averaged TPW values to the distributions of dimensionless precipitation curves. In this study, an adaptation of the

Huff curves construction methodology (Huff, 1967) based on PERSIANN data (Nguyen et al., 2019) was used for the Los Mochis study site (Figure 27).

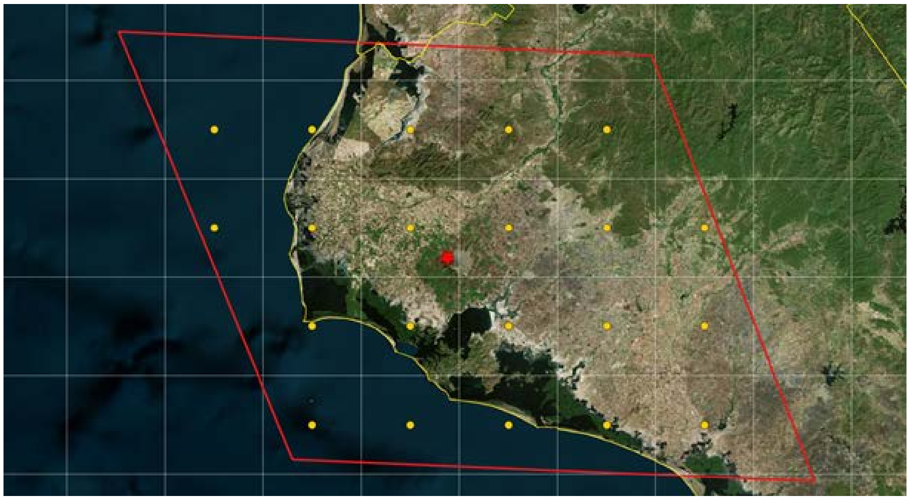


Figure 27. Los Mochis study site, the red boundary represents the ABI-GOES domain and the grid corresponds to the spatial domain of the PERSIANN database (0.25° x 0.25°). The yellow dots correspond to the PERSIANN cells used to generate the dimensionless precipitation curves, and the red dot represents the location of Los Mochis-25116 weather station (CLICOM, 2016).

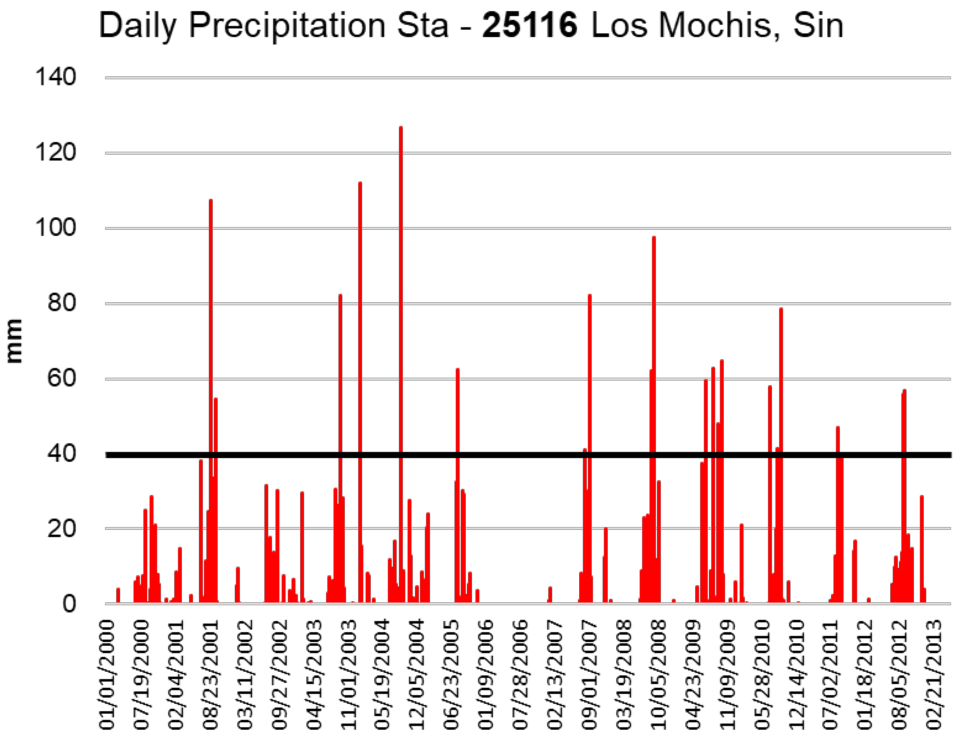


Figure 28. Daily precipitation records between January 2000 and February 2013 for Los Mochis, test site. Source: Created with information from CLICOM (2016).

A quantitative analysis of the daily precipitation records collected by Los Mochis - 25116 climate station (CLICOM, 2016) was conducted with the objective of identifying storms. All days between January 2000 and February 2013 when the daily precipitation exceeded 40 mm per day were selected as storm days and used to build a storm database (Figure 28).

Hourly rainfall information from the PERSIANN database was then used to characterize each of the storms that occurred on storm days in the absence of continuous and high-quality rainfall intensity records in this particular area. Following the methodology for constructing Huff curves (Bonta, 2004), a set of isopleths was constructed for 10, 50, and 90 % probability. This period was selected based on the availability of PERISIANN information and rainfall data collected at the Los Mochis station.

Finally, rainfall scenarios were generated by fitting the dimensionless rain mass curves (in this case, with 50% probability) with the TPW estimates obtained from the simulation of the Tropical Depression 19-E event (Dávila Ortiz, 2019).

It is important to note that a more comprehensive analysis of the precipitation patterns in the region is required in order to select an optimal MDPD value that allows the identification of storms. The main objective of this section is to demonstrate the potential for integration within the real-time flood risk estimation framework described in the previous section.

4.4. Methodology: River flooding risk assessment

4.4.1. Runoff scenarios generation

To set a precedent for the flood risk assessment complained of in phase 3 of the compressive framework proposed in this study, the precipitation scenario estimated with a dimensionless precipitation curve and TPW estimation was used to generate a runoff scenario following the methodology of a Dimensionless Unit Hydrograph

(DUH; Wanielista & Yousef, 1992), which provides insights into the river behavior in this rainfall scenario.

The DUH, a hydrologic tool developed by SCS, is widely used to estimate a watershed's runoff response to a unit pulse of precipitation. It represents the relationship between the input rainfall and the resulting direct runoff, enabling the prediction of storm hydrographs. The SCS unit hydrograph is characterized by its dimensionless nature (*i.e.*, it is independent of watershed size or storm magnitude).

The estimate of DUH requires the peak discharge value (Q_p) and the time to peak (t_p) (Equations 20 and 17). However, in this variant, instead of using a triangular hydrograph, it is fitted to a set of coordinates that emulate the distribution of a flood hydrograph obtained from field observations recorded by the SCS (Figure 29).

$$t_r = 0.6 t_c \qquad (Eq. 16)$$

$$t_p = \frac{D}{2} t_r \qquad (Eq. 17)$$

$$t_b = 2.67 t_p \qquad (Eq. 18)$$

where D is the duration of net precipitation (hours), t_r, t_c, and t_b are the lag time (hours), catchment concertation time (hours) and the base time (hours), respectively.

t / t_p	Q / Q_p	t / t_p	Q / Q_p
0,0	0	1,4	0,75
0,1	0,015	1,5	0,65
0,2	0,075	1,6	0,57
0,3	0,16	1,8	0,43
0,4	0,28	2,0	0,32
0,5	0,43	2,2	0,24
0,6	0,60	2,4	0,18
0,7	0,77	2,6	0,13
0,8	0,89	2,8	0,098
0,9	0,97	3,0	0,075
1,0	1,00	3,5	0,036
1,1	0,98	4,0	0,018
1,2	0,92	4,5	0,009
1,3	0,84	5,0	0,004

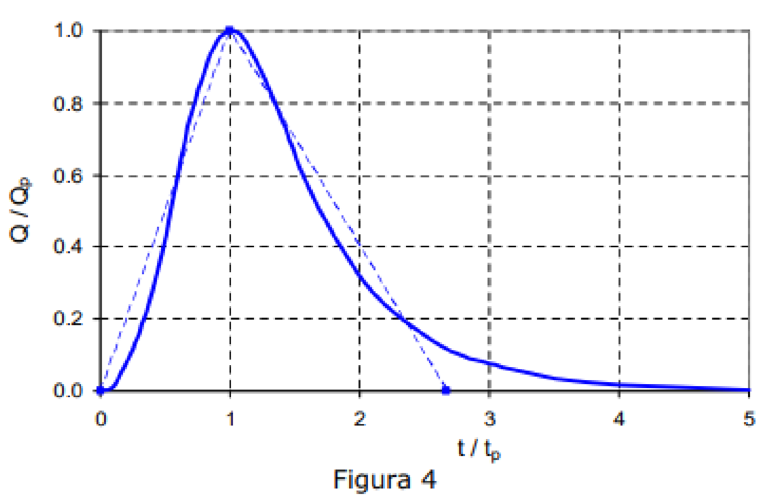


Figure 29. Coordinates of the Dimensionless Unit Hydrograph. Source: Wanielista & Yousef (1992).

According to the SCS, part of the precipitation captured in the basin does not become runoff due to losses of various types (evaporation, infiltration), so it is necessary to separate the net precipitation (P_n) from the total precipitation (P), which corresponds to this part of the precipitation that can be converted into runoff. Equations 19 and the initial value of the runoff threshold (P_0), taken from the appendix of the hydrological tables of the Ministerio de FOMENTO (2016), are used to estimate this value.

$$P_n = \frac{(P - P_0)^2}{P + 4P_0} \qquad (Eq. 19)$$

$$Q_p = \frac{P_n \cdot A_c}{1.8 \cdot t_b} \qquad (Eq. 20)$$

where P, P_n, and P_0 are expressed in mm and Q_p in m^3/s. A_c is the surface area of the basin in km^2.

Figure 30. RH10Fb basin (green) and ABI-GOES simulation domain for Los Mochis test site (red).

Runoff scenarios were generated for the RH10Fb basin (Bahía Lechuguilla-Ohuira-Navachiste; Figure 30), described in Dávila Ortiz (2019). Table 6 presents the basin parameters used to calculate the discharge and peak time, which were extracted from the INEGI watershed flow simulator (SIATL-INEGI, 2016).

Table 6. Basin parameters for RH10Fb basin. Source: SIATL-INEGI (2016).

Parameter	RH10Fb
Surface area	2,233.49 km^2
Concentration time	23.05 hours

Chapter 5. Results and Discussion

5.1. Feature importance analysis

Figures 31a– b and 31c– d present the relative feature importance identified by the RF and LR models for Los Mochis and Mexico City, respectively. In both study areas, PCA revealed that five principal components collectively explain 95% of the total variance in the datasets. This finding suggests that the dimensionality of the datasets can be reduced to five components while retaining a significant amount of information from the original dataset.

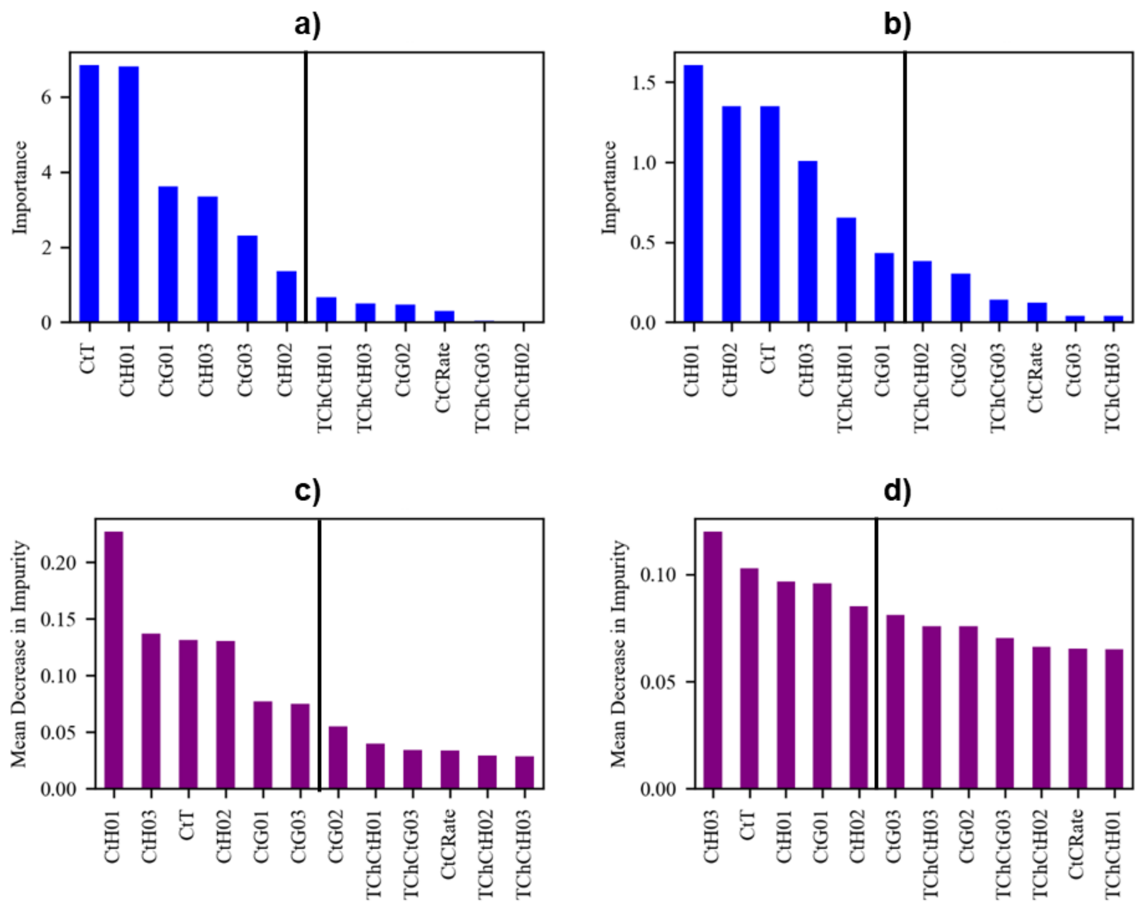


Figure 31. Feature importance analysis using the absolute value of the coefficients obtained in the training stage of the LR model (a-b) and the MDI estimated for the 12 Interest Fields with the RF model (c-d). a-c corresponds to Los Mochis and b-d to Mexico City datasets. A high MDI value and the weight of the coefficients are interpreted as indicators of the degree of contribution of a variable. The vertical line divides the interest Fields of used to construct the new (reduced) dataset.

This study selected the first five Interest Fields with the highest weighting coefficients in LR and the highest Mean Decrease in Impurity (MDI) values in RF, as outlined in Table 7. Remarkably, the Interest Fields with the most significant contributions to CC detection consistently included four predictor variables: CtT, CtH01, CtH02, and CtH03.

Table 7. Interest Fields were used as predictor variables after the feature importance analysis, for Los Mochis and Mexico City datasets.

Los Mochis	Mexico City
CtT	CtT
CtH01	CtH01
CtH02	CtH02
CtH03	CtH03
CtG01	CtG01
CtG03	TChCtH01

Consequently, the original datasets, initially characterized by 12 components, were reduced to 6, leveraging the most influential Interest Fields in the LR and RF models. This reduction in dimensionality underscores the capacity to capture critical information while enhancing the efficiency of subsequent analyses.

The recurrence with which some Interest Fields appear in both models and zones indicates the predictors that provide the most significant degree of information for deep convective cloud detection. For example, the CtT Interest Fields, which are associated with the cloud top brightness temperature values detected by the ABI-GOES 14 channel (11.2 µm TB), consistently contribute significantly at both study sites. This prominence can be attributed to the characteristic properties of deep convective clouds, where the temperature at the top of the cloud is cooler than the surrounding environment.

During the convection process, such deep clouds ascend to significant altitudes, reaching regions of lower atmospheric pressure. This ascent triggers the adiabatic cooling process, where the rising air expands and cools as it moves upward through the atmosphere, resulting in a significantly lower temperature at the top of the cloud. CtT is an Interest Field with a high degree of importance that recurs in other works.

For example, D. Han et al. (2019), S. Lee et al. (2017), and Mecikalski et al. (2015) reported CtT as a significant contributor to the identification of Convective Initiation events (*i.e.,* the probability that a given cumulus cloud object will develop into a ≥35-dBZ-intensity radar echo at –10°C altitude; Gravelle et al., 2016). Furthermore, in Kim et al. (2017), CtT was identified as the most influential variable in Overshooting convective cloud Tops (OTs) classification by both RF and LR models.

The following Interest Fields, in order of importance, are CtH01, CtH02, and CtH03. These fields, with their spectral differences between ABI-GOES channels, provide valuable insight into the height and depth of the cloud-top (S. Lee et al., 2017). Similar to CtT, CtH01 emerges as one of the variables with a higher degree of contribution in CC class detection. This Interest Field is used to determine the lower stratospheric moisture and shows positivity when water vapor is detected above cloud tops, a key indicator of the presence of OTs formations (Kim et al., 2017). On the other hand, for the glaciation indicators, only CtG01 was identified as one of the most significant variables, although ice production in the upper levels of the cloud occurs regularly near the time of maximum cloud growth (Henderson et al., 2021).

Despite the inclusion of the temporal trend variables as Interest Fields designed to convey information about the rate of vertical cloud-top growth, these predictors generally contributed relatively less in both the LR and RF models (except for TChCtH01 at the Mexico City site). This finding highlights the necessity for further research and refinement in this field.

Figures 32 and 33 illustrate the distribution of selected fields of interest after feature importance analysis for Los Mochis and Mexico City, respectively. In the Los Mochis dataset, the CC and noCC classes show statistically discernible differences in the six selected Interest Fields, thus improving the classification task in the ML models. There are minimal Interquartile Ranges (IQRs) for the CC class, with a median close to 0 for spectral difference fields. It is consistent with the claim of Kim et al. (2017) that when a cloud reaches its local equilibrium height or the height of the tropopause, all channel differences tend to approach zero.

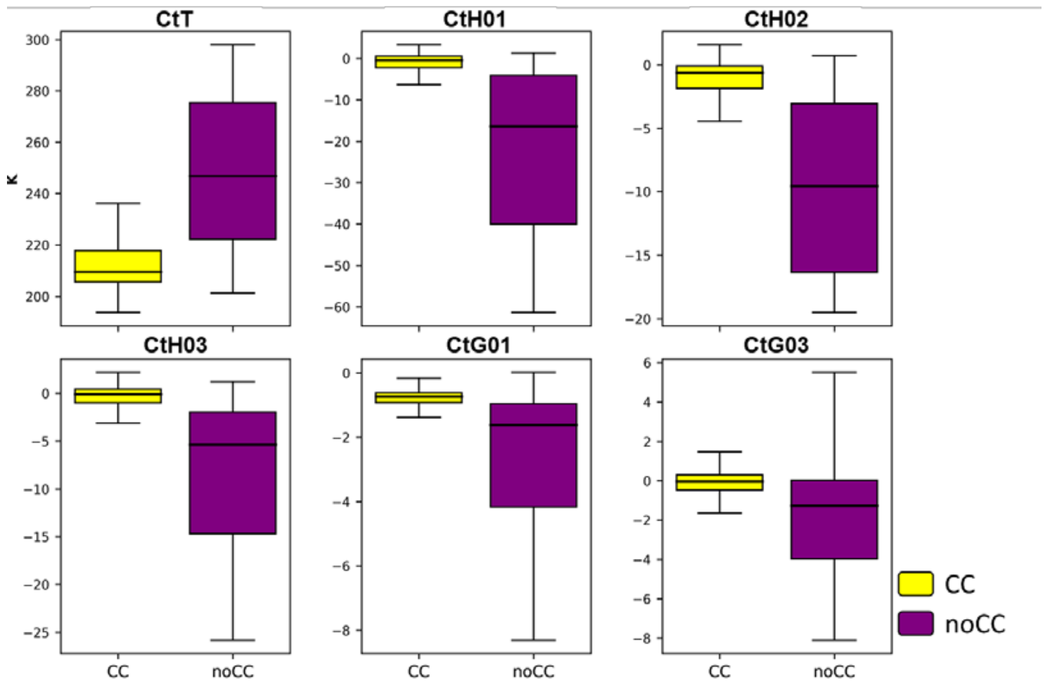


Figure 32. Box plots of input variables for reference data used in the CC detection models for Los Mochis. The predictive dataset was generated from the most significant features of the LR and RF models. The number of classes labeled as CC is 24,116 and noCC is 10,884.

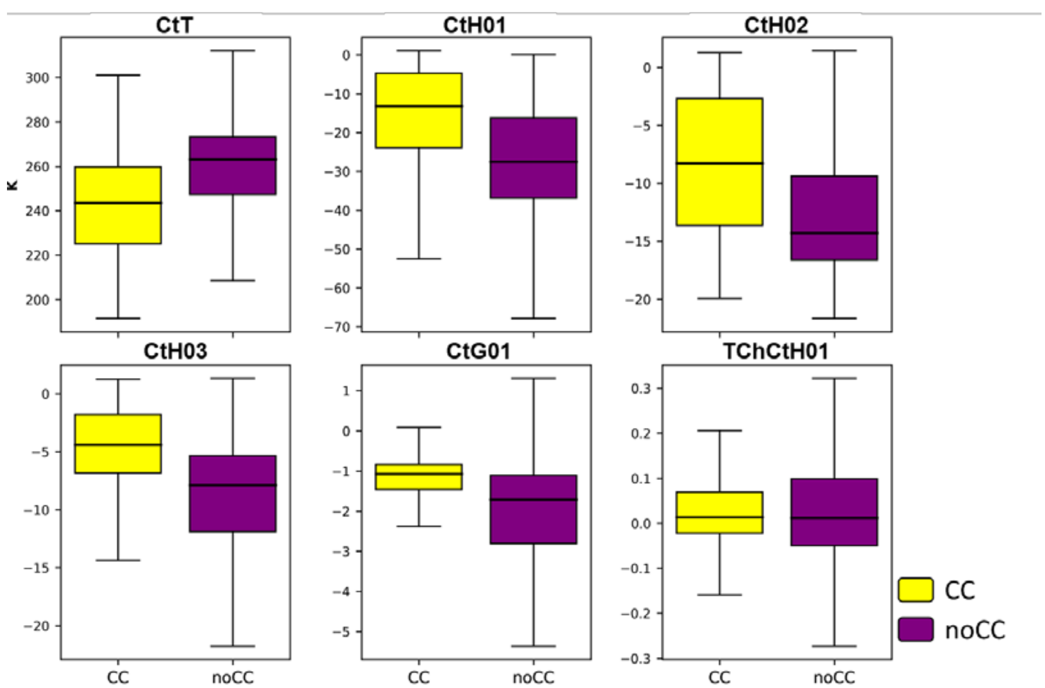


Figure 33. Box plots of input variables for reference data used in the CC detection models for Mexico City. The predictive dataset was generated from the most significant features of the LR and RF models. The number of classes labeled as CC is 57,833 and noCC is 79,667.

In contrast, the data distribution for the noCC class is much broader. In the Cloud-top Temperature (CtT) field, cells present higher values because the tops of deep convective clouds are colder due to adiabatic cooling. In the spectral difference fields, these are mostly negative values.

For the Mexico City dataset, overlapping distributions can be seen in the boxplots of the selected Interest Fields. It is reflected in the performance metrics of the ML models, which notably present lower values than the metrics observed in the Los Mochis case. However, the overall trends are preserved, and the CtT values are lower for the CC class. In addition, the noCC values were lower in the height and glaciation spectral difference fields. Regarding the IQRs of the predictors, the distributions appear more balanced for both classes, which, together with the distribution overlap, allows us to infer greater complexity in modeling CC events at Mexico City than at Los Mochis.

The CtT Interest Field's significant contribution to detecting the CC class is evident. Cells with lower temperature values are associated with potential storm clouds. In this context, da Silva Neto et al. (2016) list some cloud-top temperature thresholds reported in studies related to the identification of deep convective clouds, such as 241 and 221 K (Maddox, 1980), 221 K (Anderson & Arritt, 1998), 225 K (K. Bedka et al., 2010), 255 and 206 K (Machado & Rossow, 1993). Additionally, Siqueira et al. (2021) analyzed 7139 mesoscale convective events over southeastern Brazil and estimated mean brightness temperature values of 224 K at convective initiation, 220 K at maturation, and 225 K at decay. This research estimated that storm clouds manifest at a CtT threshold of <220 K for the Los Mochis study site and <260 K for Mexico City.

5.2. Hyper-parameters Tuning results

Figure 34 shows the results of the hyperparameter tuning process for the LR, RF, and MLP models using the cross-validation strategy. These parameters control the complexity and regularization of the ML models and, therefore, need to be specified in advance. Additionally, these parameters can significantly influence the performance of the ML models (Ukkonen & Mäkelä, 2019).

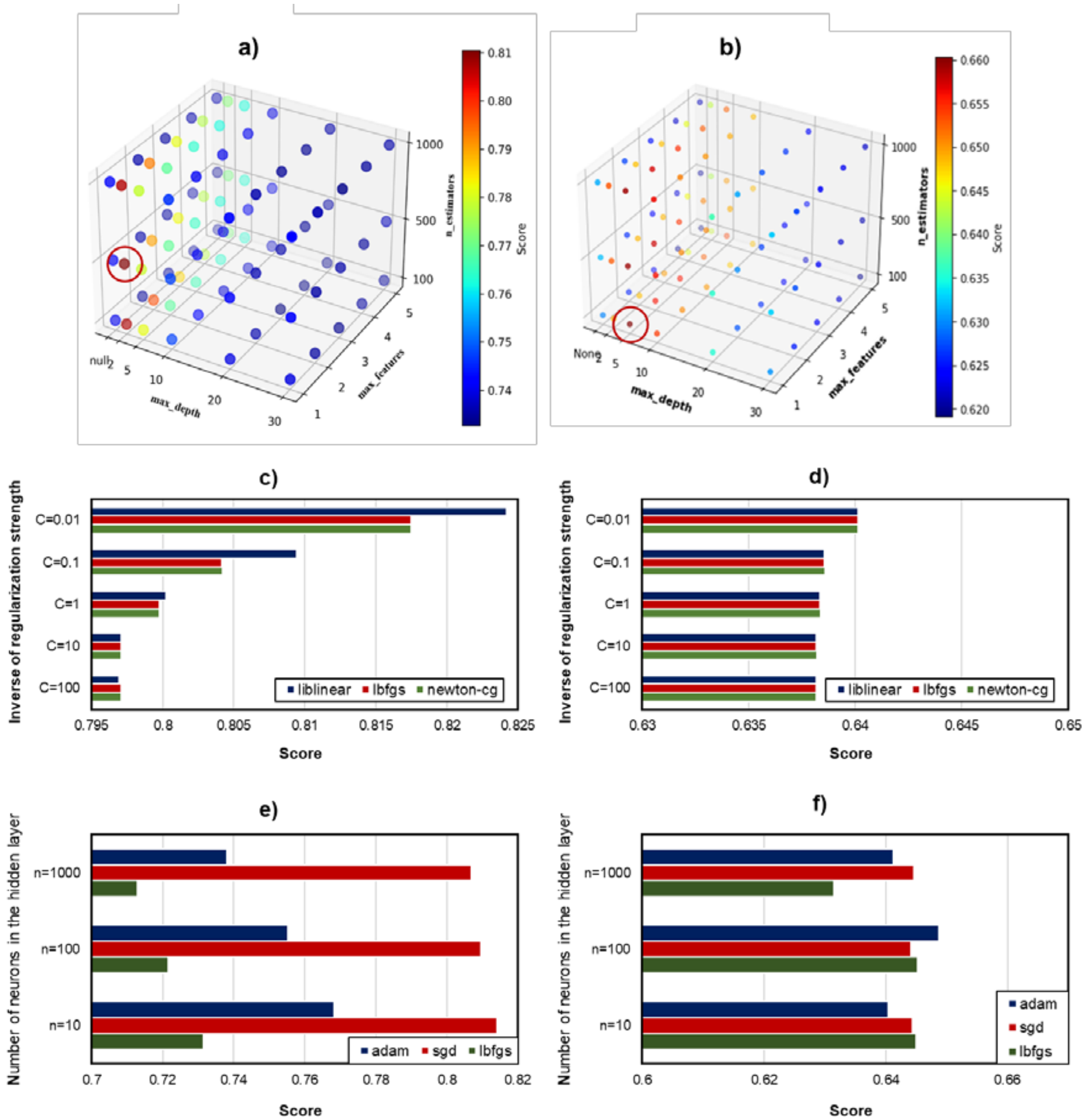


Figure 34. Hyperparameter tuning results for RF (a, b), LR (c, d), and MLP (e, f). The plots on the left correspond to Los Mochis dataset (a, c, e), and those on the right correspond to the Mexico City dataset (b, d, f).

Regarding LR, the highest accuracy values for both study areas were obtained with a configuration of C = 0.01 and the L2 penalty rule (Table 8). The parameter C is the inverse of the regularization strength, with smaller values of C corresponding to more robust regularization. Regularization prevents overfitting by penalizing large coefficient values, thus favoring simpler models. When C is large, the model minimizes the misclassification of training data, potentially leading to complex models that fit noise. On the other hand, smaller C values emphasize coefficient regularization, favoring models that may generalize better to new data (Dávila Ortiz et al., 2023a). Similarly, L2 helps prevent overfitting by discouraging overly complex models with large coefficient values. This regularization term is proportional to the square of the size of the coefficients. The liblinear and Newton-cg solvers are numerical optimization algorithms which find optimal coefficients that minimize the logistic loss function.

Table 8. Summary of the hyperparameters of LR, RF, and, MLP were tuned using the cross-validation strategy.

Hyperparameter	Los Mochis	Mexico City
Norm of the penalty (LR)	L2	L2
The inverse of regularization strength (LR)	0.01	0.01
Solver (LR)	liblinear	newton-cg
Maximum depth of the tree (RF)	2	5
Number of trees in the forest (RF)	500	100
Random split predictor variables (RF)	1	1
Number of neurons in the hidden layer (MLP)	10	100
Solver (MLP)	SGD	Adam

For RF, there was a significant difference between Los Mochis and Mexico City datasets regarding the number of trees in the forest (number of estimators), which was 500 and 100, respectively. In the context of RF, this parameter determines the number of decision trees (DT) included in the ensemble. The maximum depth of the trees is higher in the Mexico City dataset, which could be because, at a greater depth, each tree can capture more complex patterns in the data.

The number of neurons (n) in the hidden layer of an MLP neural network is a flexible parameter that adapts to the complexity of the dataset and the nature of the underlying patterns. In this sense, the Mexico City dataset (n=100) may contain more intricate and complex patterns than the Los Mochis dataset (n=10), necessitating a more expressive model to learn the intricate features and representations in the data.

Stochastic Gradient Descent (SGD) and Adam are optimization algorithms used to update the neural network weights during the training process. SGD works by updating the model parameters (weights) in a direction that reduces the loss function for a randomly selected subset of the training data (mini-batch), where each mini-batch is randomly selected from the training set. Adam is an adaptive learning rate method that efficiently updates the model weights during training. This algorithm helps accelerate convergence, especially in varied and complex data scenarios.

5.3. Performance and validation of detection models

Each ML approach was evaluated using six performance metrics (POD, FAR, ACC, BIAS, CSI, and F1) that indicate each model's ability to detect convective cells, false alarm rate, etc. The effect of integrating a post-processing filter lightning incidence-based was also evaluated.

Figures 35 and 36 show the results for datasets corresponding to Los Mochis and Mexico City, respectively. For the Los Mochis dataset, there was a marginal difference between the results of the ML approaches (ML series) and those with the integrated post-processing filter (ML+LF). In this regard, the ability of the models to detect the CC class, represented by the POD metric, presents values close to 0.8. At the same time, the FAR value for all ML approaches remains below 0.2. This means that potential profound convective events can be detected proficiently without including lightning incidence data. However, this does not imply that lightning is infrequent in the region; on the contrary, Holle & Murphy (2015) reported a high density of lightning strikes in northeastern Mexico. The region with the highest density of lightning strikes is in the narrow strip between the SMO and the Gulf of California, with peaks observed in July and August, coinciding with the dates used to construct the data sets for this research. In addition, GLM-GOES lightning strikes were identified in 8 of the 14 images used to construct the training and test datasets.

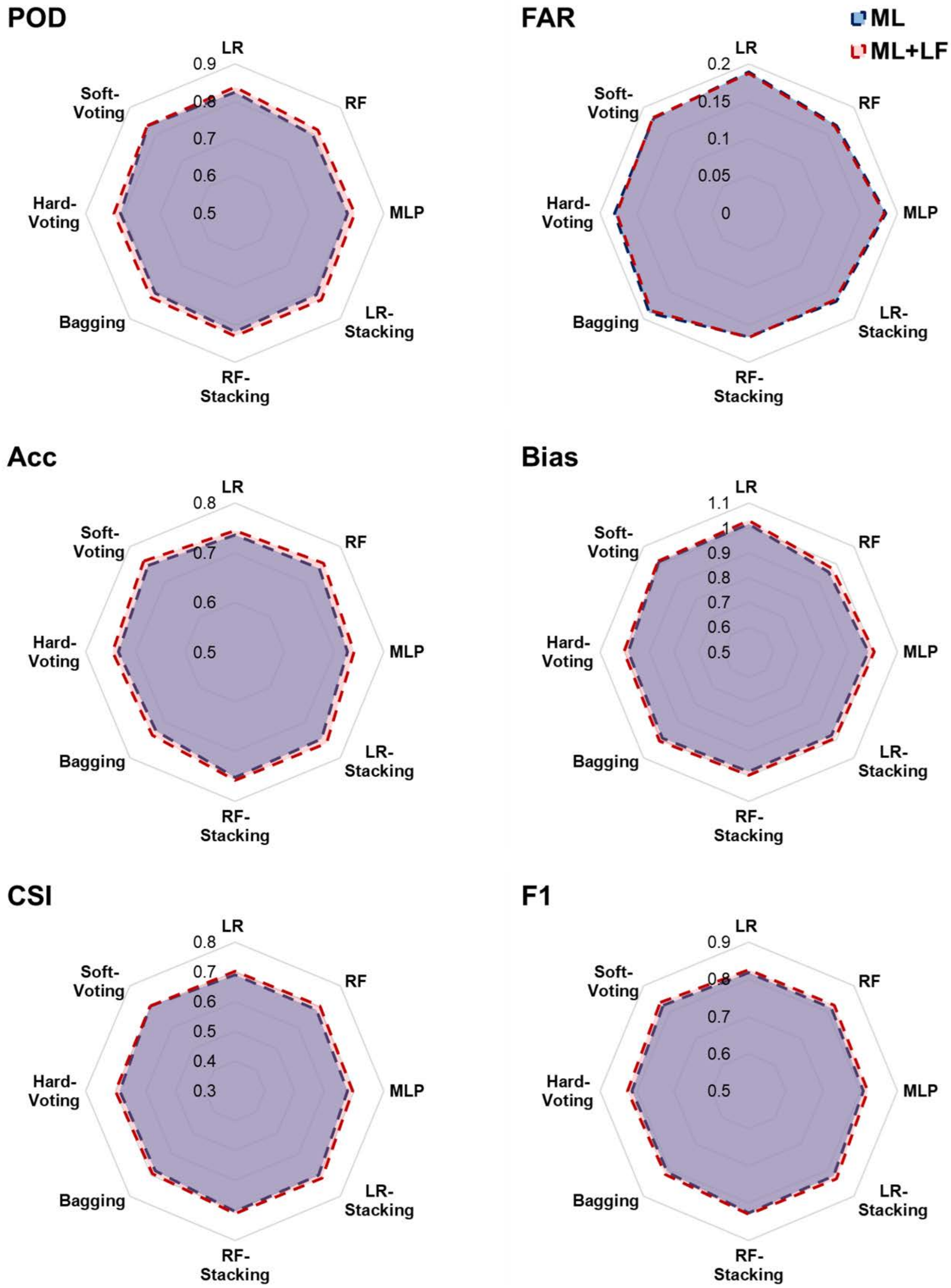


Figure 35. Evaluation results of the POD, FAR, ACC, BIAS, CSI, and F1 metrics for the ML approaches using Los Mochis test dataset. Each vertex of the graph corresponds to an ML approach, the blue lines correspond to the performance results of the ML models (ML), and the red lines show the results with Lightning Filter integration (ML+LF).

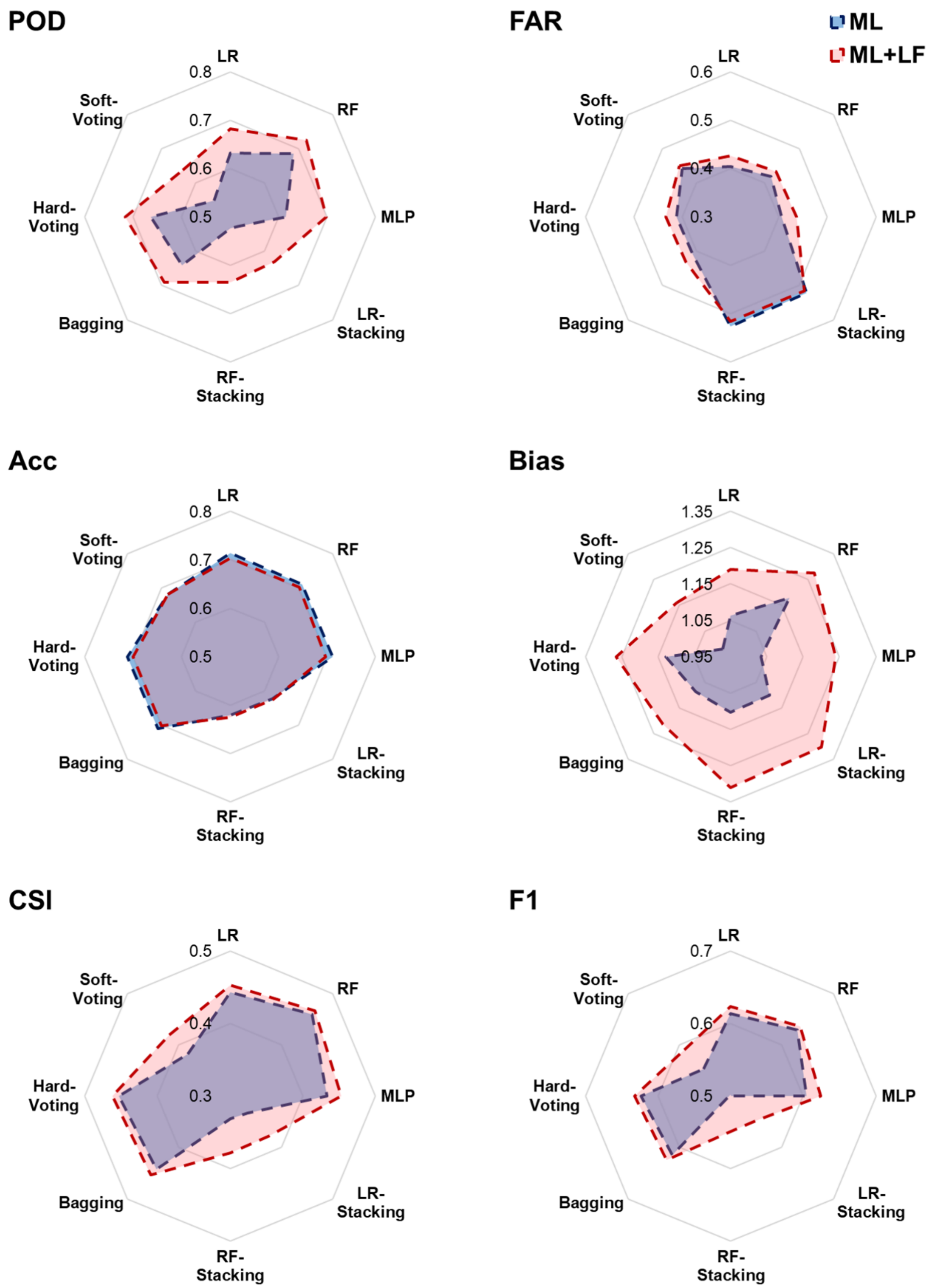


Figure 36. Evaluation results of the POD, FAR, ACC, BIAS, CSI, and F1 metrics for the ML approaches using Mexico City test dataset. Each vertex of the graph corresponds to an ML approach, the blue lines correspond to the performance results of the ML models (ML), and the red lines show the results with Lightning Filter integration (ML+LF).

Alternatively, the Mexico City site shows significant improvement when the Lightning Filter is integrated, mainly showcasing the POD, CSI, and F1 metrics. For example, the highest value was POD = 0.68, obtained using RF, which increased to POD = 0.72 after the post-processing filter. The second-highest POD was obtained using the LR, which increased from 0.63 to 0.7 after integrating the LF. Contrary to the Los Mochis dataset, for Mexico City, the integration of GLM data significantly enhances the ability of the system to detect the CC class across all ML approaches. This improvement is attributed to the fact that the occurrence of lightning strikes is a reliable indicator of the generation of a deep convective event, which is consistent with the intense lightning activity in the region.

In addition to POD, both CSI and F1 metrics show improvements in ML approaches with initially lower overall performance, such as LR-stacking and Soft-Voting for Mexico City dataset. In contrast, the Lightning Filter negatively affects the Bias metric, increasing its values beyond the optimal 1. This metric, which ranges from 0 to infinity, assesses whether the forecasting method tends to underestimate (BIAS<1) or overestimate (BIAS>1) the CC class (Siqueira et al., 2021). In this context, the presenting Bias values greater than 1 with the Lightning Filter integration may indicate a limitation in the labeling of convective cells with the MODIS sensor, which has a lower resolution than the ABI-GOES products. This is because there are areas labeled as noCC that present lightning activity that indicates the generation of a convective environment. Future work will address the integration of unsupervised learning approaches to simplify the generation of a reference set using a higher-resolution sensor.

The analysis of the Los Mochis dataset revealed no significant differences between the use of model ensembles and more straightforward approaches such as LR. This finding is consistent with the results of previous studies (Kim et al., 2017; Mecikalski et al., 2015; Ukkonen & Mäkelä, 2019) that reported this model as a robust alternative for convective hazard modeling and detection. In contrast, the performance of each ML approach differed significantly for the Mexico City dataset. The results demonstrated that LR-stacking, RF-stacking, and Soft-Voting exhibited inferior performance for the majority of the evaluated metrics. Conversely, LR, RF, MLP, Bagging, and Hard-Voting consistently demonstrated comparable and satisfactory performance across all six metrics.

Regarding the POD values, the acceptable threshold may vary based on the specific forecasting or detection system goals. However, a high POD value may be necessary to minimize the risk of missing hazardous events. In this sense, for the Los Mochis dataset, the highest value was POD = 0.84, while, in the Mexico City dataset, it was POD = 0.72; both obtained LR after post-processing filtering. In this context, Kim et al. (2017) and Kim et al. (2018), in their work on OT detection, report performances of POD = 0.77 and POD = 0.79, respectively, while S. Lee et al. (2017) and H. Han et al. (2015) in their studies on CI detection estimated values of POD ≈ 0.8 and POD > 0.7. Examples from the literature, such as Liu et al. (2019), with their proposed convective environment alerting and monitoring system, obtained a POD between 0.66 and 0.7; on the other hand, Zhang et al. (2019), with their convective storm nowcasting based on CNN, achieved POD values close to 0.7.

The FAR metric provides information about the classifier making incorrect positive predictions. In the case of Los Mochis, the FAR values were below 0.2, with minimal variation between the different ML approaches, with values between 0.16 and 0.18. In Mexico City, significant variation was observed, with LR-stacking and RF-stacking having the highest FAR values at 0.51, indicating a higher incidence of false alarms. Other ML approaches, particularly LR and RF, also showed increased FAR values, but to a lesser extent. In this case, LR changed from FAR = 0.40 to 0.42, while RF changed from FAR = 0.41 to 0.43 in the ML+LF series. Compared with the existing literature, the FAR values vary significantly depending on the type of convective forecast performed. For example, Kim et al. (2017) and Kim et al. (2018) reported FAR values between 0.3 and 0.09, while K. M. Bedka & Khlopenkov (2016) obtained FAR values between 0.01 and 0.18. For CI detection, H. Han et al. (2015) estimated FAR values between 0.46 and 0.83, S. Lee et al. (2017) a FAR ≈ 0.2, in Mecikalski et al. (2015) the FAR value ranges between 0.22 and 0.36, and finally in D. Han et al. (2019) it ranges from 0.46 to 0.37.

Accuracy (Acc) provides an overview of the effectiveness of a model in correctly classifying samples. In the Los Mochis dataset, there are no significant differences between the Acc values of the ML approaches, which are generally slightly higher than 0.7. In this case, adding LF leads to a marginal increase in the Acc value, although not significant. Conversely, in Mexico City, the lowest values were recorded for LR-Stacking and RF-Stacking, with Acc = 0.62, while LR, RF, Bagging, and Hard-Voting had Acc values of 0.7. However, this metric is susceptible to bias in datasets with class imbalance because it does not consider the distribution of classes in the dataset. Therefore, it is imperative to complement the assessment of model performance with other metrics. The similarity in Acc values between Los Mochis and Mexico City is noteworthy, especially considering the higher class imbalance in Los Mochis (CC class = 69%, noCC class = 31%) compared to the Mexico City dataset (CC class = 42%, noCC class = 58%).

Bias is another standard metric in studies related to convective event detection (*e.g.*, Siqueira et al., 2021), as well as in convective hazard forecasting such as lightning (*e.g.*, Zhou et al., 2020), as it assesses whether the forecast method tends to underestimate or overestimate CC events. The estimated Bias values for Los Mochis range from 0.97 to 1.03, indicating a relatively acceptable performance in terms of false alarms and misses. In contrast, significant variation is observed between ML approaches in Mexico City, with values exceeding Bias > 1.15 for RF and escalating to Bias > 1.25 after LF.

The comprehensive analysis of different metrics facilitates a thorough evaluation of the forecasting framework proposed in this study. Thus, integrating GLM data increases the CC class's overestimation rate concerning the target reference dataset. However, it enhances the system's ability to detect CC events.

CSI assesses a forecast's accuracy in predicting a specific event's occurrence (*e.g.*, severe thunderstorms) relative to actual observations. It is particularly useful in situations where false alarms or missed events can have important consequences, such as severe weather forecasting. For the Los Mochis dataset, CSI values close to 0.7 were estimated, implying that a CSI greater than 0.7 indicates a robust classification model. The model accurately identifies positive cases while minimizing false alarms and misses. The CSI results for the Mexico City dataset range between 0.4 and 0.5, slightly increasing after post-processing with LF, especially for the worst-performing metrics such as LR-Stacking, RF-Stacking, and Soft-Voting. A CSI value between 0.4 and 0.7 indicates moderate success in the classification task. Although the model accurately predicts positive instances, there is room for improvement in reducing false positives.

Regarding the F1 metric, a high F1 score, closer to 1, indicates a model with high precision and recall. In the context of the Los Mochis, it can be inferred that all ML approaches effectively classify positive instances while avoiding false positives and false negatives. Similarly, the post-processing LF improves the F1 score for models with lower performance in the Mexico City dataset. For specific ML approaches, such as LR, RF, MLP, Bagging, and Hard-Voting, the F1 values are estimated to be F1 = 0.63.

Figure 37 shows the Receiver Operating Characteristic (ROC) curves for the ML approaches for the Los Mochis and Mexico City datasets. These curves showcase the trade-off between sensitivity and specificity at different threshold values. A classifier with a curve closer to the upper-left corner indicates better performance by achieving high accurate positive rates while maintaining low false positive rates. For the Los Mochis dataset, a generally homogeneous behavior was observed, except for RF-Stacking, which showed a lower discriminative capacity. The area under the ROC curve (AUC) values is higher than 0.7, indicating that the ML approaches generally have an acceptable discrimination capacity between classes. In the case of Mexico City, the ROC curve analysis revealed that LR, RF, MLP, Bagging, and Hard-Voting consistently outperformed over several thresholds. These classifiers showed ROC curves with a steeper ascent, suggesting enhanced discrimination between CC and noCC cases. This finding implies that these ML approaches are well-suited for CC detection in this specific dataset. Nonetheless, it is essential to consider the insights provided by other metrics in conjunction with the ROC analysis.

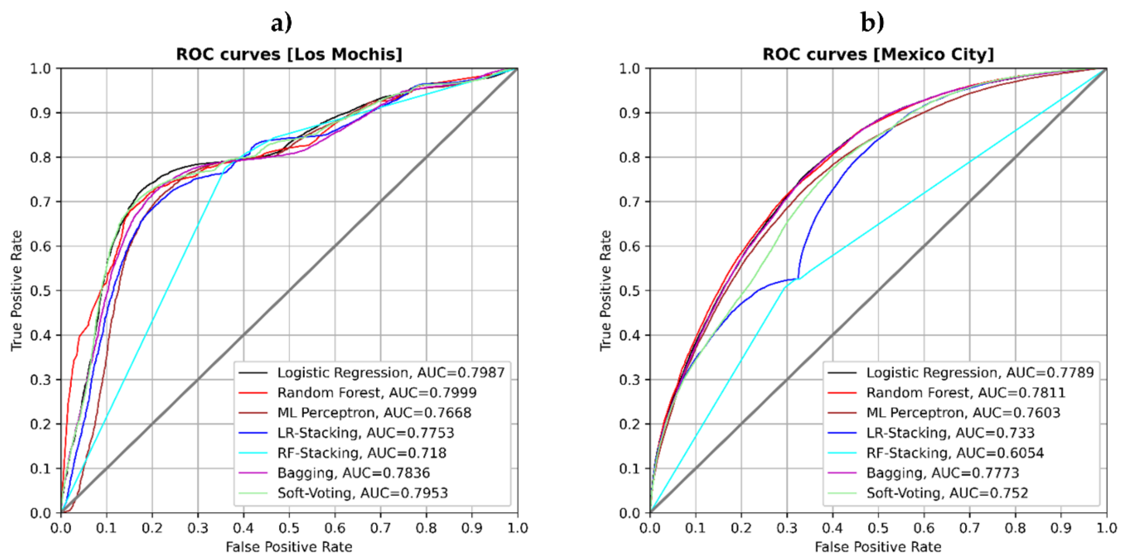


Figure 37. ROC curves and AUC values for the predictions obtained with LR (black), RF (red), MLP (brown), LR-Stacking (blue), RF-Stacking (cyan), Bagging (magenta), and Soft-Voting (light green) for Los Mochis (a) and Mexico City (b) datasets. These ROC curves provide a comparison of the sensitivity and specificity of each ML approach at different discrimination thresholds.

5.4. Simulation of reference convective events

Examples of deep convection events simulated from each ML approach for Los Mochis and Mexico City are shown in Figures 38-41. This comparison allows a qualitative evaluation of the performance of each model, while the IoU metric indicates the overlap between the predicted regions and the ground truth (Table 9).

Table 9. IoU metric results for the Los Mochis 2018-227 - 18:00 (August 15, 2018), Los Mochis 2019-234 - 17:30 (August 22, 2019), Mexico City 2019-247 - 20:10 (September 04, 2019), and Mexico City 2021-135 - 19:50 (May 15, 2021) events for each ML Approach. Note: These events were taken from the test data set for each study area.

Event	LR	RF	MLP	Bagging	LR stacking	RF stacking	Hard voting	Soft voting
Los Mochis 2018-227 - 18:00	0.86	0.76	0.56	0.67	0.44	0.62	0.65	0.65
Los Mochis 2019-234 - 17:30	0.86	0.84	0.67	0.72	0.57	0.59	0.77	0.77
Mexico City 2019-247 - 20:10	0.53	0.51	0.51	0.53	0.41	0.42	0.52	0.45
Mexico City 2021-135 - 19:50	0.54	0.54	0.54	0.54	0.34	0.34	0.55	0.41

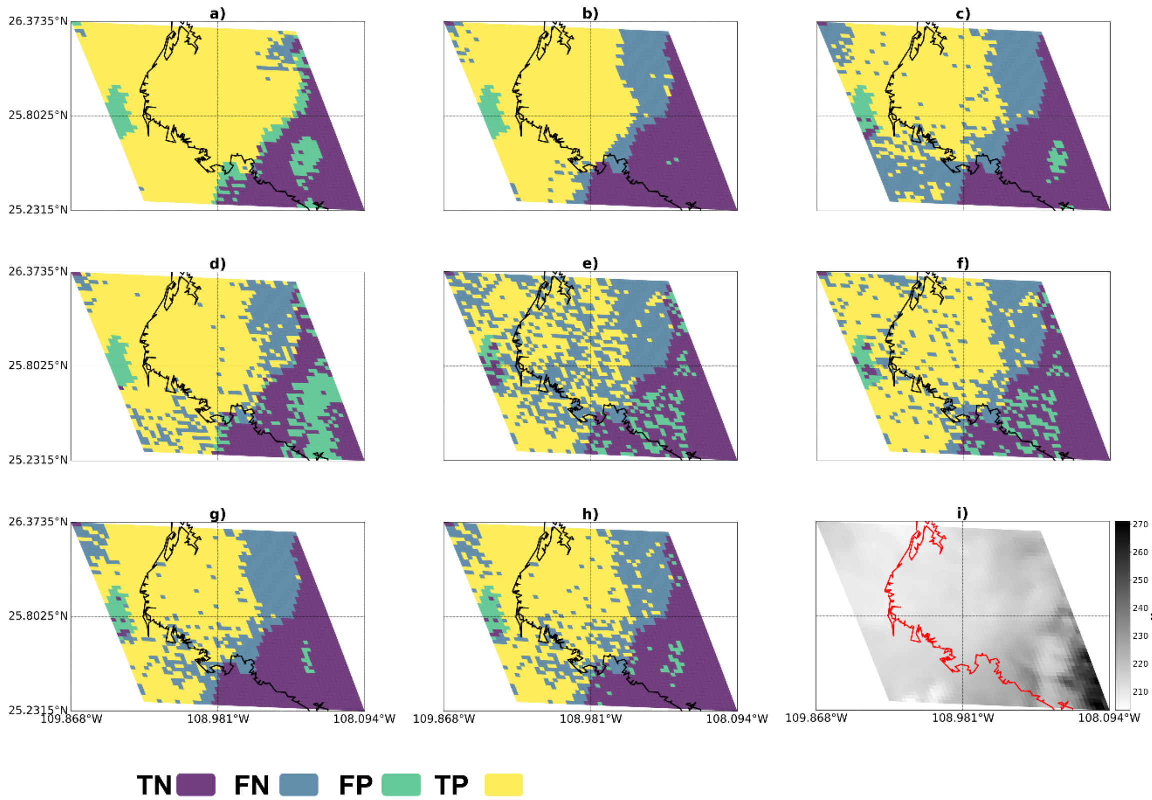


Figure 38. IoU maps of the Los Mochis 2018-227 - 18:00 (August 15, 2018) event, generated with the reference labels and simulations of the eight ML approaches LR (a), RF (b), MLP (c), Bagging (d), LRstacking (e), RFstacking (f), Hvoting (g), Svoting (h), and Cloud-top brightness temperature map of the same event obtained from the ABI-GOES channel 14 (i). In general, most ML models underestimate the cloud boundary, except for LR, where the identification of a potential unlabeled convective core is simulated. The remnants of this core are also presented in MLP, Bagging, and Svoting. Salt-and-pepper noise is observed in LRstacking and RFstacking. TN: True Negative values, FN: False Negative values, FP: False Positive values, and TP: True Positive values.

In the Los Mochis dataset, the simulations of the August 15, 2018 (Figure 38) and August 22, 2019 (Figure 39) events showed a tendency to underestimate the total area of deep convective clouds at their boundaries, except for the LR model, where the highest IoU values were estimated (IoU = 0.86 in both cases).

The presence of false negative pixels at the cloud edge indicates the difficulty in accurately modeling this transition zone, where cloud properties derived from ABI-GOES data become diffuse (Dávila-Ortiz et al., 2024). Comparing the event simulations with the Cloud-top temperature maps, it can be seen that there is a good correspondence between the spatial distribution of the CC pixels and the coldest areas of the images. However, in both events, these colder regions present a larger area in the simulated CC class and target labels, adding uncertainty about the actual extent of the deep convection event. Therefore, future work will explore adding a transition class between CC and noCC.

In Figure 38, most ML approaches, excluding RF, simulate a convective core in the lower right part of the simulation domain, which is inconsistent with the reference labels for this event. This discrepancy could be attributed to the better ability to detect deep convection events from AGI-GOES data compared with MODIS.

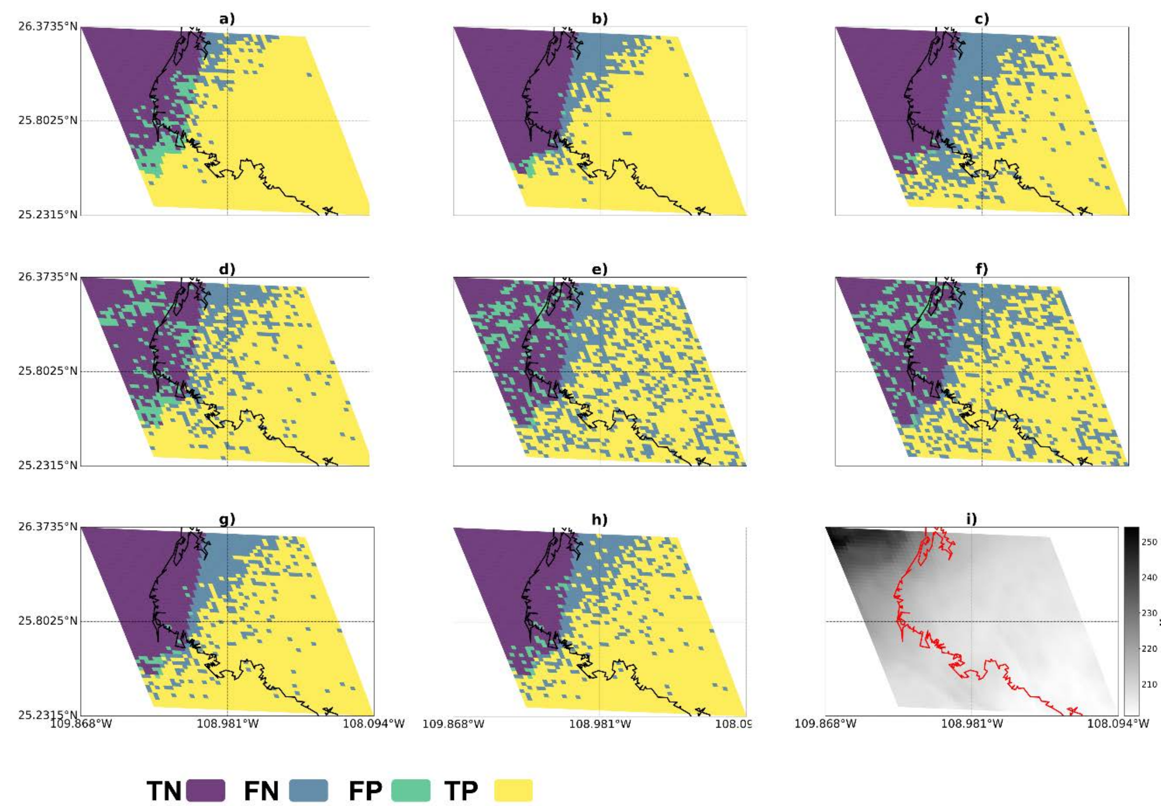


Figure 39. IoU maps of the Los Mochis 2019-234 - 17:30 (August 22, 2019), event generated with the reference labels and simulations of the eight ML approaches LR (a), RF (b), MLP (c), Bagging (d), LRstacking (e), RFstacking (f), Hvoting (g), Svoting (h), and Cloud-top brightness temperature map of the same event obtained from the ABI-GOES channel 14 (i). The majority of the ML models showed the tendency to underestimate the CC class at the edge of the convective cloud and the presence of salt-and-pepper noise in MLP, Bagging, LRstacking and RFstacking and Svoting.

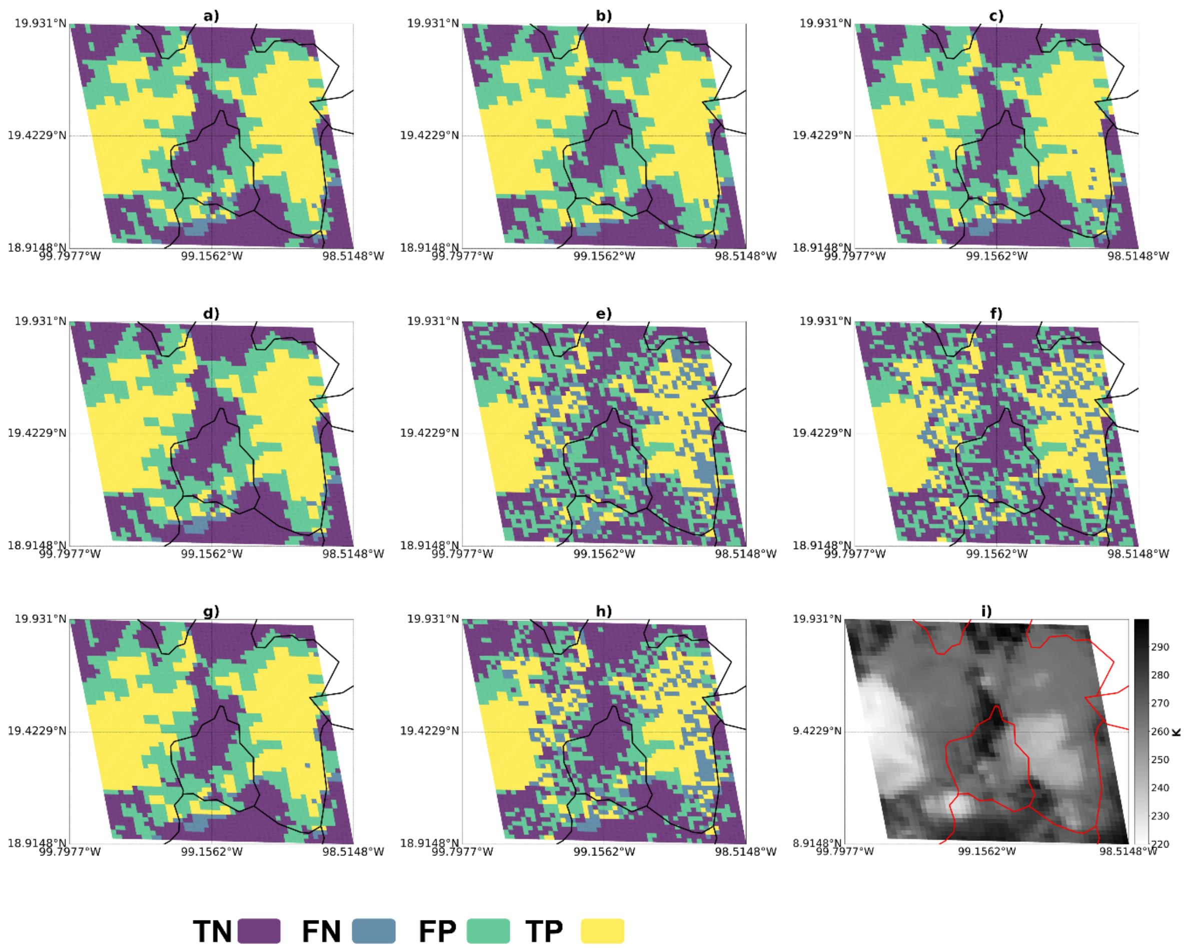


Figure 40. IoU maps of the Mexico City 2019-247 - 20:10 (September 04, 2019), event generated with the reference labels and simulations of the eight ML approaches LR (a), RF (b), MLP (c), Bagging (d), LRstacking (e), RFstacking (f), Hvoting (g), Svoting (h), and Cloud-top brightness temperature map of the same event obtained from the ABI-GOES channel 14 (i). The outputs of LR, RF, MLP, Bagging, and Hvoting are consistent with each other and show a significant level of overestimation at the edge of the cloud. On the other hand, the outputs generated by LRstacking, RFstacking, and Svoting are extremely noisy, especially at the boundary of the clouds.

The ensemble models LRstacking and RFstacking show the worst performance, with IoU values of 0.44 and 0.62, respectively, for the 2018-227 - 18:00 event and IoU values of 0.57 and 0.59 for the 2019-234 - 17:30 event. These ML approaches observe a pronounced salt-and-pepper effect when no contextual spatial information is provided (Kete et al., 2019).

This pattern is also evident in the simulated events for Mexico City (see Figures 40 and 41), where LRstacking, RFstacking, and Svoting similarly demonstrate this effect, which is strongly correlated with the low-performance metrics of the test dataset. In the event Mexico City 2019-247 - 20:10 (Figure 40), two well-separated clouds can be seen forming over regions of higher topographic elevation, together with the presence of cores of lower cloud top temperature, allowing the inference that their formation was due to a forced convection process.

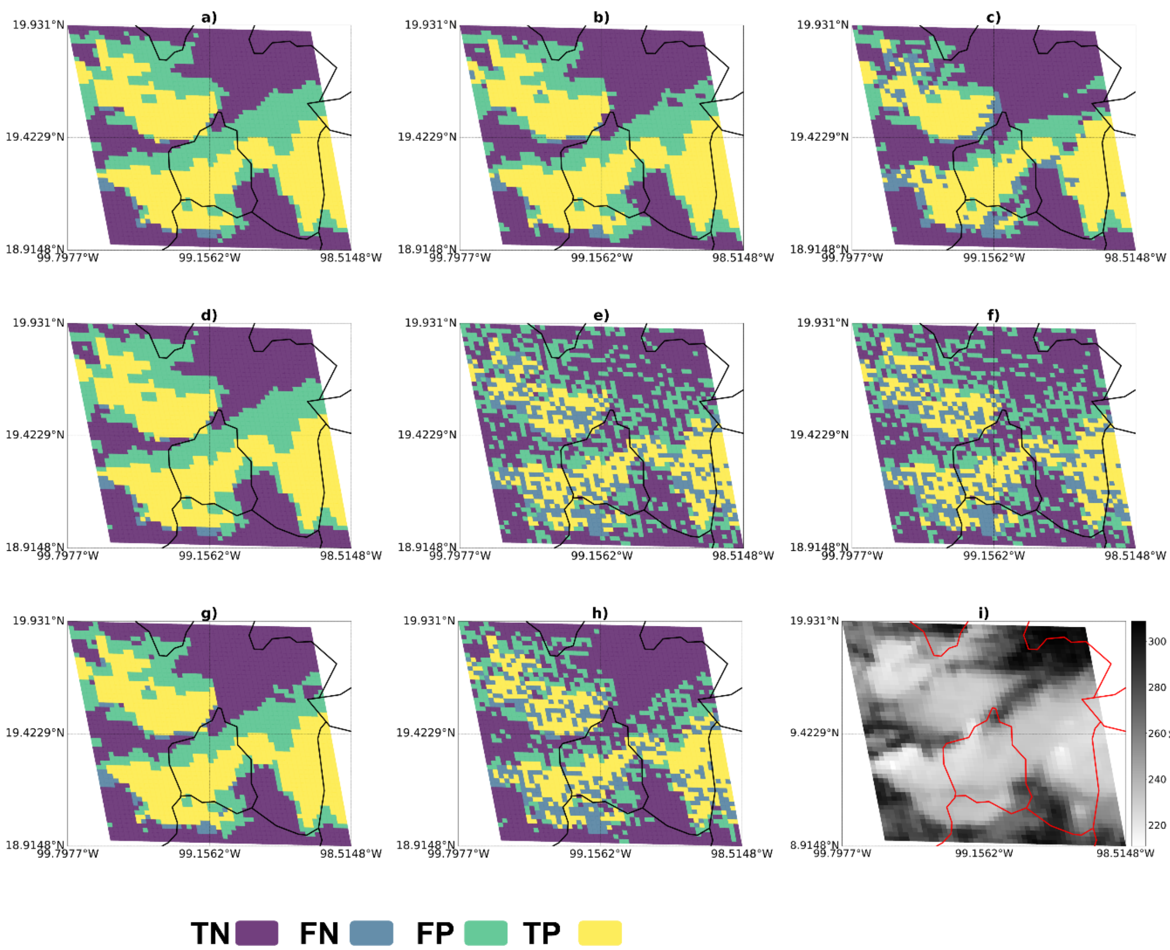


Figure 41. IoU maps of the Mexico City 2021-135 - 19:50 (May 15, 2021), event generated with the reference labels and simulations of the eight ML approaches LR (a), RF (b), MLP (c), Bagging (d), LRstacking (e), RFstacking (f), Hvoting (g), Svoting (h), and Cloud-top brightness temperature map of the same event obtained from the ABI-GOES channel 14 (i). The simulations obtained with LRstacking, RFstacking, and Svoting are particularly noisy; however, the remaining ML approaches correctly simulate deep convection zones with a significant degree of underestimation at the edges of these regions.

In the case of the two events simulated for Los Mochis, all ML approaches demonstrated limitations in accurately modeling the cloud boundary. This may be attributed to a possible diffusion effect of the optical and microphysical properties at the boundaries of the deep convective cloud or to inherent constraints of the labeling process with MODIS data. A similar situation occurs in the case of Mexico City, where the area of both convective events is overestimated for this spatial domain. Furthermore, there is a consistent presence of false positive pixels in the cloud edges (Figures 40 and 41). However, these overestimation zones are consistent with cooler zones mapped by the ABI-GOES Channel 14. As in the previous cases, addressing this area of improvement could entail the addition of a transition class or the adoption of alternative methods that preserve the spatial contextual structure, such as Convolutional Neural Networks (CNN) or other spatially aware models. For example, object-based forecasting models offer the advantages of identifying pertinent regions for severe weather prediction and reducing the volume of data processed (Gagne et al., 2017).

5.5. Computational costs of ML approaches

Figure 42 comprehensively evaluates the computational costs associated with the eight ML approaches in both Los Mochis (Figure 42a) and Mexico City (Figure 42b). This assessment, crucially, is conducted within a standardized computing environment, considering both computational time and memory usage. Doing so provides reliable insight into each model's efficiency and resource requirements, ensuring a fair and consistent comparison.

The results of the Los Mochis dataset showed notable variations in the computational cost across the evaluated ML approaches. RF and Bagging demonstrated efficient computation times but consumed more memory. This highlights their resource-intensive nature, primarily due to the sampling with replacement, which increases the dataset size during training. LR and MLP presented a balanced trade-off between time and memory usage, with LR generally showing low computing time because it involves simple matrix operations and low memory consumption. After all, LR only stores the coefficients of the linear model. On the other hand, in this analysis, the MLP remains a simple configuration with only one hidden layer. Ensemble models present increased memory usage in this scenario because of their greater complexity. However, this trend does not hold in the computational time domain, where LR, RF, and Bagging were the ML approaches with less time consumption, showing only marginal variation.

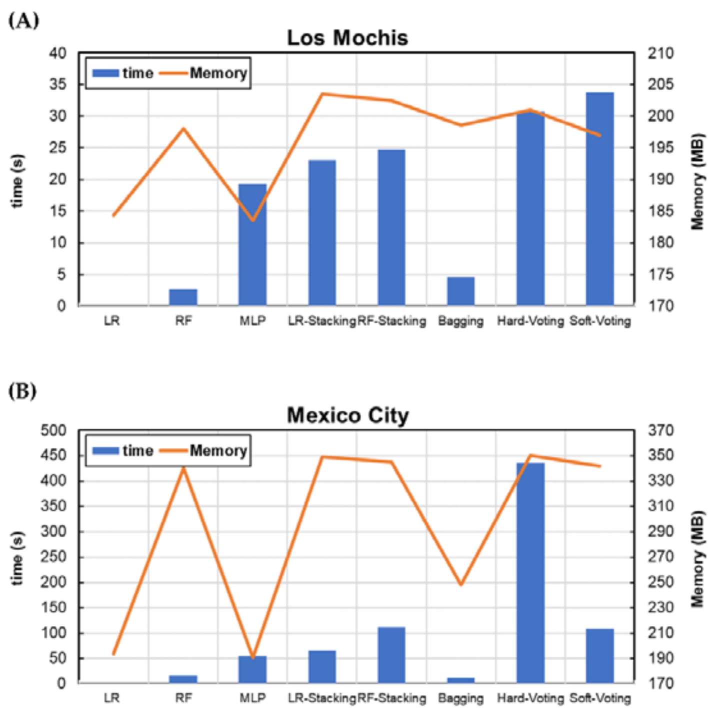


Figure 42. Computational cost graph for each ML approach corresponding to Los Mochis (a) and Mexico City (b). The right axis shows the total simulation time and the left axis shows the memory usage.

In the Mexico City dataset, Hard-Voting exhibited a longer computing time than other ML approaches. It's noteworthy to mention that the omission of SVM in Soft-Voting can reduce the processing time compared to Hard-Voting (Figure 42b). For Los Mochis, RF demonstrated efficient computation times but consumed more memory (Figure 42a). Conversely, LR, RF, and Bagging demonstrated competitive performance in both time and memory. A more comprehensive comparison should consider the specific requirements of the application, the size of the dataset, and available computing resources.

5.6. Dimensionless precipitation curves for Los Mochis, Sin

Figure 43 shows the summary process for generating 10%, 50%, and 90% curves (isopleths) of the probability distributions of dimensionless depths (Huff Curves; Figure 43d) based on Huff methodology modified by Bonta (2004) for the Los Mochis domain.

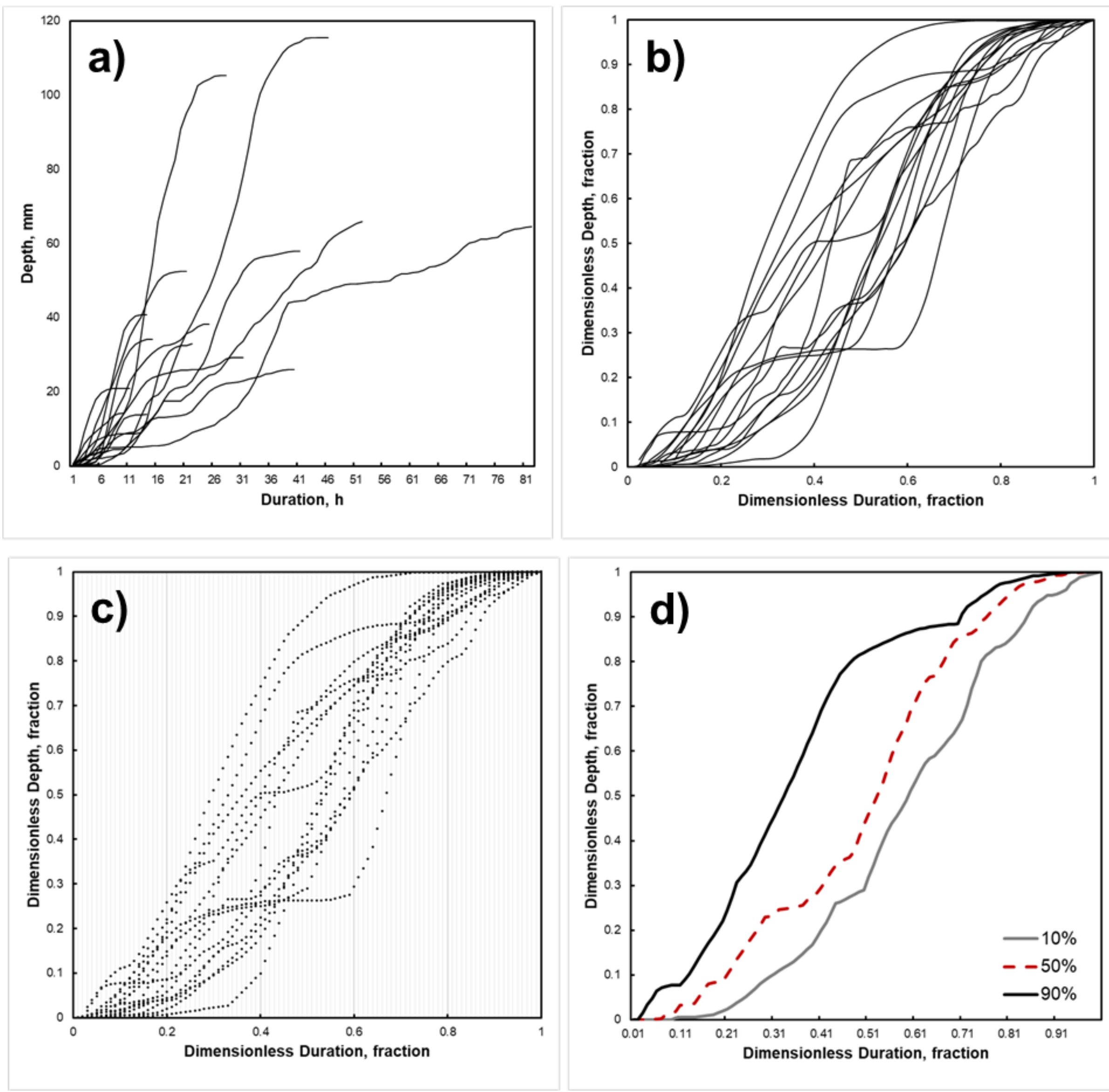


Figure 43. Step1: Storm mass curves for storms identified. Storm totals >40 mm (a). Step2: Dimensionless storm mass curves (b). Step3: Storm mass−curve intersections with Duration times (c). Step4: Isopleths of probability (Huff Curves; d).

These probability isopleths comprise a set of "Huff Curves," a probabilistic summary representation of storm mass curves in terms of dimensionless elapsed times into a storm and corresponding dimensionless accumulated depths (Figure 43d). In this context, a 90% isopleth implies that for all storm durations, 90% of the accumulated precipitation has occurred for all dimensionless storm durations (Bonta, 2004). For most purposes, median curves (50% isopleth) are probably most applicable to the design. Extreme curves (10 % or 90 %) can be helpful when runoff estimates are needed for unusual storm conditions (Elfeki et al., 2014).

The dimensionless precipitation curves for the Los Mochis spatial domain were derived from the estimated hourly precipitation data sourced from the PERSIANN database (0.25° x 0.25°; Nguyen et al., 2019). These data were collected for storm days recorded by the Los Mochis - 25116 climate station (CLICOM, 2016). The subsequent step in generating precipitation scenarios, based on TPW derived from a deep convective event, is discussed by considering the simulation of the Tropical Depression 19-E event (Dávila Ortiz, 2019).

5.7. Generation of Rainfall-Runoff scenario assessment

From the RF model, which together with LR were the ML approaches that achieved the best performance metrics in the Los Mochis dataset, the Tropical Depression 19-E event was simulated, which had significant impacts on September 19 and 20, 2018 (Dávila Ortiz, 2019). This event was selected due to the intense convective activity that occurred in various areas of northwest Mexico, leading to high-intensity storms. According to Protección Civil Municipal, over 350 mm of rainfall was reported during the night of September 19 (El Debate, 2018).

Figure 44 compares the simulation of the event as it reached its highest TPW peak with the temperature at the top of the cloud, showing a particularly close relationship between the simulated CC class distribution and the presence of cooler cloud regions.

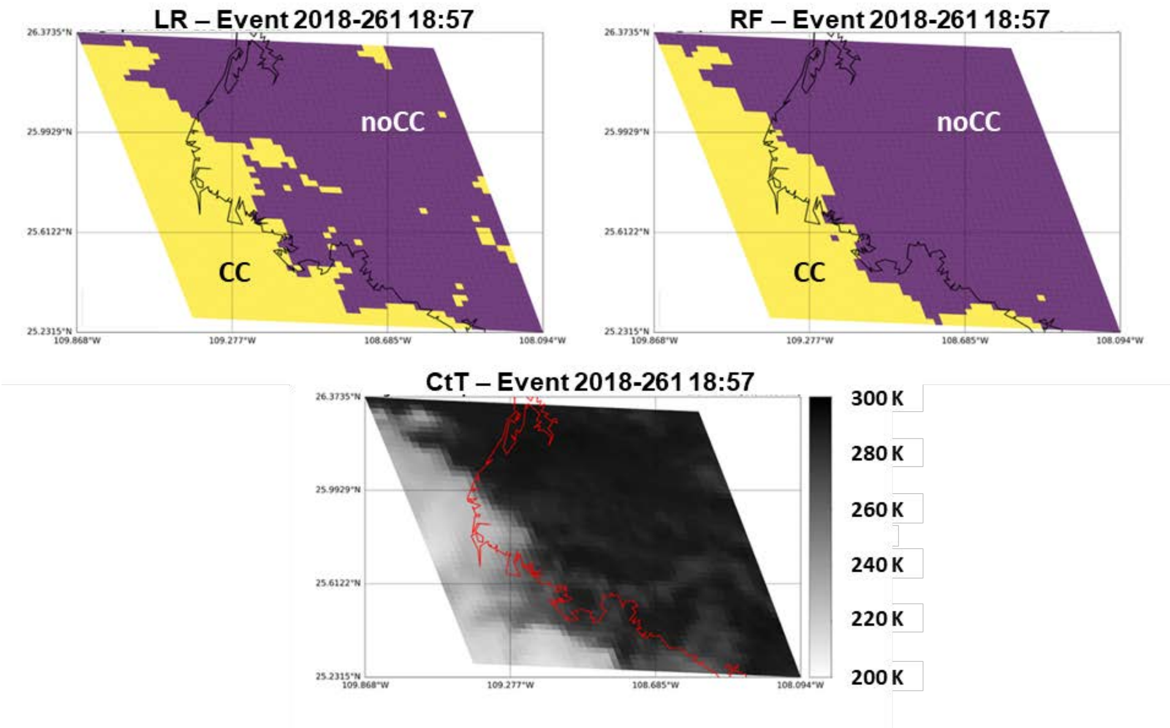


Figure 44. Comparison between RF and LR simulations with Cloud-top brightness temperature map for Los Mochis 2018-261 18:57 event (Tropical Depression 19-E).

Figure 45a shows the estimated mean TPW values for all simulated CC pixels between each satellite view of the 19-E event over the Los Mochis domain (Figure 1), each 5 min. Although these estimates represent the average TPW values of the entire simulation domain, which may consist of one or more convective cores or several independent vertically developing clouds in the mature phase, this value is a good indicator of the level of risk from the impact of a potential storm within the spatial domain of the simulation. In this sense, it is noteworthy that the TPW estimates for this event range between 50 and 60 mm from an early stage of event formation; this does not mean that all regions of the domain simulated with the CC class are homogeneous; therefore, it could be inferred that certain areas could present even higher estimated TPW values. This could be related to the high rainfall intensity during this severe storm.

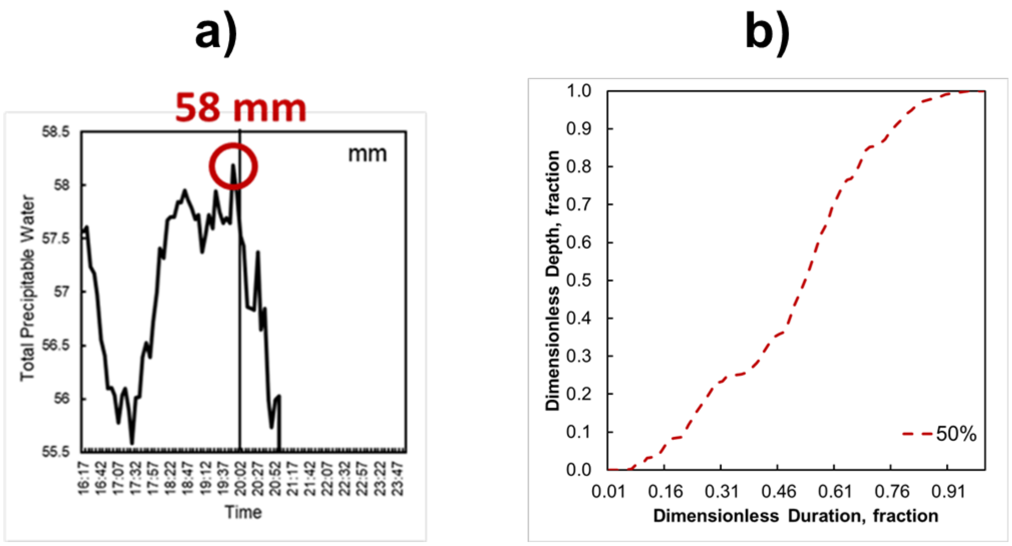


Figure 45. a) Estimated average TPW values using detected CC pixels as the extraction mask array for each satellite view of the 19-E event over the Los Mochis domain (each 5 min). b) 50% Isopleths of probability.

Once the TPW values derived from the ABI-GOES products are extracted, they are averaged, and adjusted to the distribution of the isopleths generated with information from the Los Mochis spatial domain. For this example, we chose rainfall with a duration of 6 h and a 50% curve representing the median of the analyzed storms (Figure 45b). However, these parameters can be modified according to the objective of the rainfall scenario and the forecaster's criteria.

The result of disaggregating the peak TPW value of 58 mm is the precipitation scenario (hyetogram) shown in Figure 46. This scenario, adjusted to a total depth of 58 mm and a total duration of 6 h (360 min), corresponds to the 18:45 UTC of the 19-E event. Once the presence of the CC class is detected in the spatial domain and the TPW values from ABI-GOES products are available, this scenario can be updated according to the time distribution of the system proposed in this work. This update can occur every 5 minutes, aligning with the availability of information from the ABI-GOES sensor.

Finally, Figure 47 shows the runoff scenario derived from the precipitation scenario of the 19-E event and the hydrographic parameters of the RH10Fb basin. From the TPW estimation during the analyzed convective event, a peak flow of up to 250 m/s^3

was estimated, indicating a significant risk level due to main stem overflow, floodplain inundation, presence of urban flash floods, and landslide hazards.

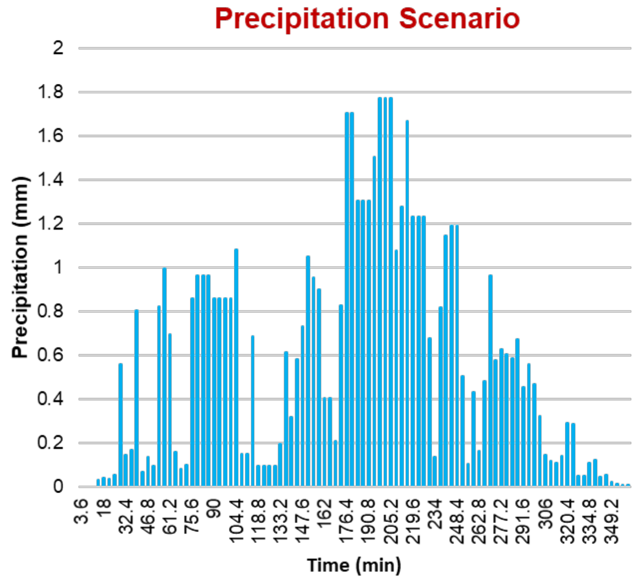


Figure 46. Precipitation scenario derived from the average TPW of simulated CC pixels in Los Mochis 2018-261 18:57 event (Tropical Depression 19-E) and the 50% Huff isopleth. The hyetograph has a duration of 6 h and a total precipitation depth of 58 mm.

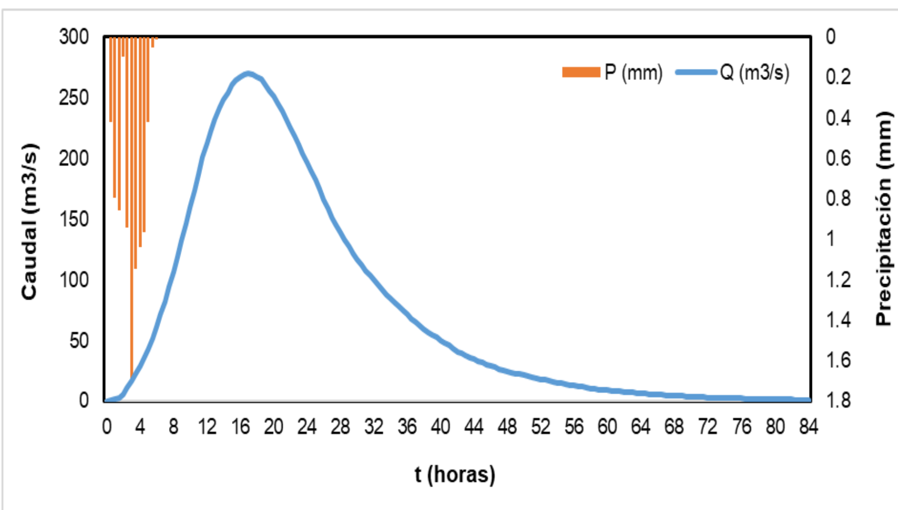


Figure 47. Estimated hydrograph (Runoff scenario) for the RH10Fb basin, based on the precipitation scenario generated for Los Mochis 2018-261 18:57 event (Tropical Depression 19-E).

The river flood estimation analysis based on the runoff scenario provides valuable insights into the potential impact of extreme rainfall events on flooding within the main river basin.

Chapter 6. Example of application: Buoyancy force analysis

6.1. Deep convective cloud formation events

One of the principal axes of this research, in addition to developing and implementing an EWS in the face of deep convective events, is its application as a tool for studying and analyzing the formation processes and physical properties of storm clouds in different regions of Mexico.

In this context, 12 intense convective events were collected (Table 10) in five different spatial domains in Mexico (Figure 48). These domains include the Los Mochis domain (Figure 1) and an extension of the Mexico City domain (Figure 3) to cover a broader area of the Trans-Mexican Volcanic Belt. These events were selected using an alternative criterion to delineate potential areas with the presence of deep convective clouds, adapting the ISCCP methodology (Rossow & Schiffer, 1991) of simple thresholding, cells with COD (Cloud Optical Depth, instead of Cloud Optical Thickness; Table 2) values greater than 23 and CtP values less than 440 mb are referred to as deep convective clouds (Figure 11).

Table 10. List of events included in this example of application.

Name	Spatial domain	Date
Event 01	Trans-Mexican Volcanic Belt section	Jun 17, 2018
Event 02	Trans-Mexican Volcanic Belt section	Sep 18, 2018
Event 03	Trans-Mexican Volcanic Belt section	Sep 16, 2021
Event 04	Chiapas-Tabasco	May 02, 2019
Event 05	Chiapas-Tabasco	Oct 15, 2022
Event 06	Los Mochis, Sin	Jul 17, 2018
Event 07	Los Mochis, Sin	Sep 18, 2018
Event 08	Ciudad Victoria, Tamps	May 21, 2021
Event 09	Ciudad Victoria, Tamps	Sep 05, 2022
Event 10	Álvarez Mountain Range	May 09, 2018
Event 11	Álvarez Mountain Range	Jun 09, 2020
Event 12	Álvarez Mountain Range	Sep 17, 2022

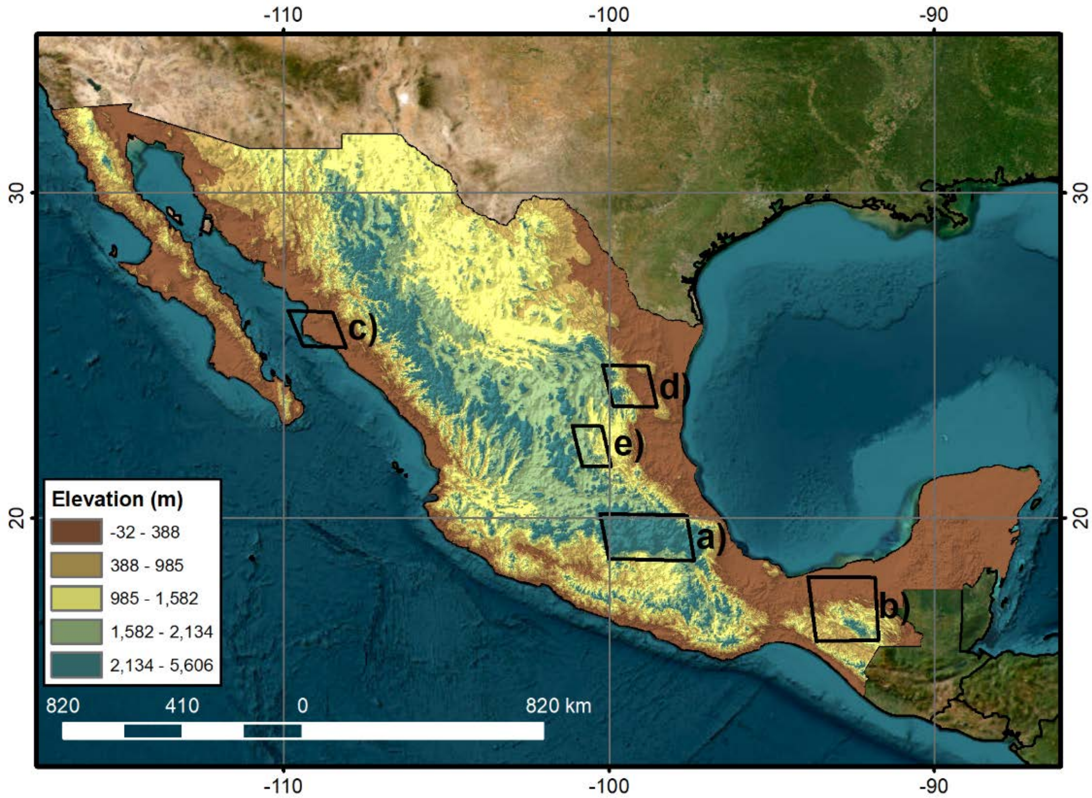


Figure 48. Selected spatial domains in this example of application. a) Trans-Mexican Volcanic Belt section, b) Chiapas-Tabasco, c) Los Mochis, Sin, d) Ciudad Victoria, Tamps, and e) Álvarez Mountain Range.

This approach has important limitations compared with the one proposed by the ML models explored in this study. First, this method is based on MODIS products; therefore, adaptation to ABI-GOES data is required. In addition, the spatial resolution of the products derived from GOES-R differs from that of the ABI-GOES bands. (Table 1). For example, the COD product image is produced on the ABI fixed grid with a spatial resolution of ~2 km for the CONUS coverage region (Figure 8) and ~10 km for the CtP product (NOAA & NASA program, 2019).

However, with the proposed method, the process becomes simpler. By selecting a thresholding criterion based on ABI-GOES derived products (Table 2), the need to consolidate a reference dataset and train ML models to detect potential deep convective clouds is eliminated. This approach, as demonstrated in the binary labeling process of Convective Cells (CC) and noCC in Figure 49, is straightforward and flexible, making it more accessible for a wider range of study areas.

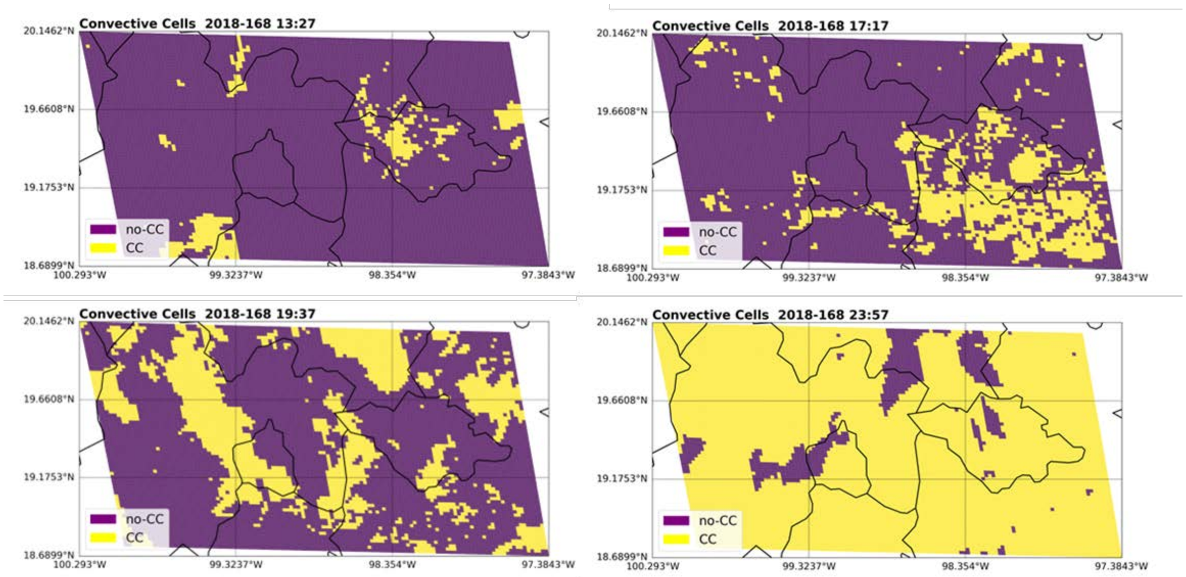


Figure 49. Binary labeling of a convective event using a thresholding approach for different time steps for Event 01 (Trans-Mexican Volcanic Belt section). Yellow cells correspond to CC and purple cells to no CC.

While the selection of different spatial domains was driven by the intense convective activity characteristic of these regions due to different convective processes, a detailed description of the physical terrain conditions as well as the atmospheric dynamics process is beyond the scope of this study.

This section only addresses the analysis of various optical and microphysical properties within the identified vertical development clouds. The goal is to explore the potential implementation of a tool for studying storm clouds based on the architecture and code developed for the proposed framework. A future study will consider a comprehensive analysis of these 12 convective events and their understanding of atmospheric cloud formation processes.

The potential of this application is then illustrated through the analysis of formation variables and cloud properties, focusing on four reference events selected from Table 10. The variables considered include CtH (Cloud-top Height), CtT (Cloud-top Temperature), CtP (Cloud-top Pressure), COD (Cloud Optical Depth), and TPW (Total Precipitable Water). These variables will be studied to gain insight into cloud formation and evolution dynamics.

6.2. Deep convective clouds properties

- Topographic Configuration

Figure 50 shows the topographic configuration within each spatial area delineated in Figure 48, corresponding to Events 02, 05, 07, and 09 (Table 10). The area corresponding to the Trans-Mexican Volcanic Belt section (Figure 50a) shows intricate topographic features related to regional tectonic activity resulting from the interaction of the North American, Pacific, and Cocos tectonic plates. León-Cruz et al. (2021) emphasize the considerable role of these complex relief features in influencing forced convection processes. On the other hand, within the Chiapas-Tabasco domain (Figure 50b), we can see a section of the Sierra Madre de Chiapas, which is part of the American Cordillera, a chain of mountain ranges that consists of an almost continuous sequence of mountain ranges that form the western of North America, Central America, and South America.

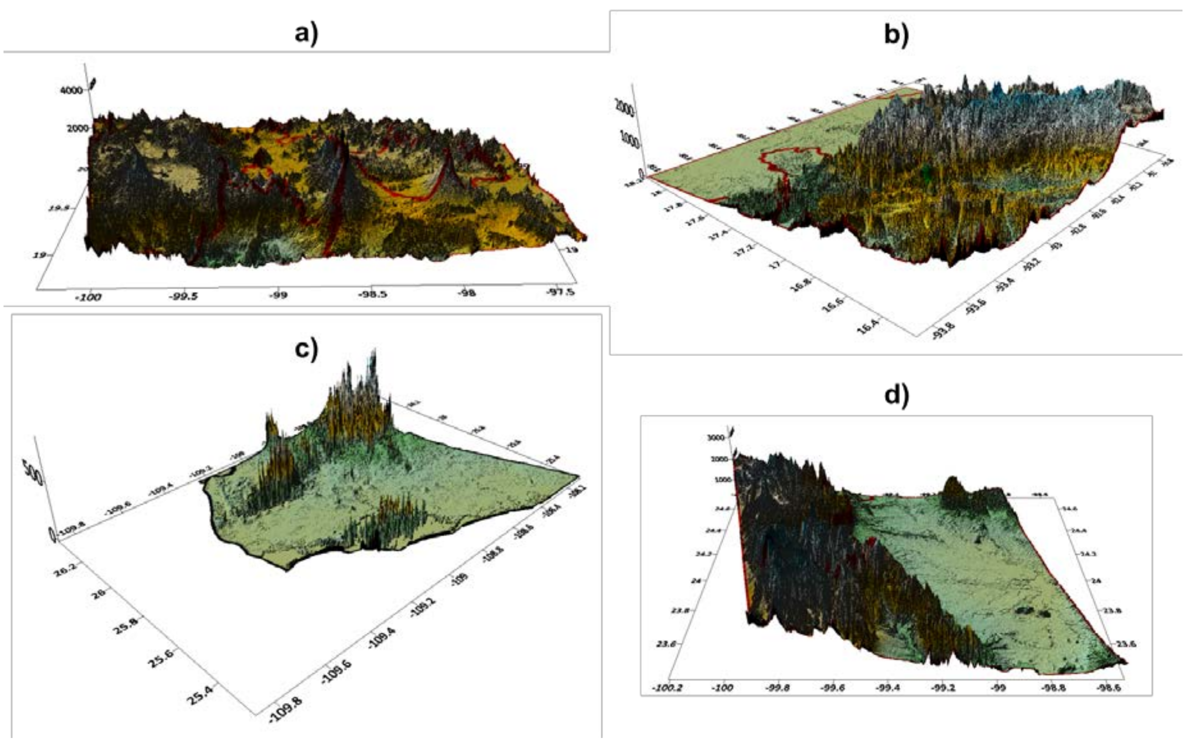


Figure 50. Relief representation of the spatial domains in 3D. The red border indicates state boundaries, while the black border indicates coastlines. a) Trans-Mexican Volcanic Belt section, b) Chiapas-Tabasco, c) Los Mochis, Sin, and d) Ciudad Victoria, Tamps. The 3D models were generated using information from INEGI (2017).

Figure 50c shows the relief features of Los Mochis, positioned between the Gulf of Mexico to the west and the SMO to the east. The terrain is predominantly flat, with some peaks over 500 meters in elevation. In contrast, the domain of Ciudad Victoria, Tamps. (Figure 50d), marks a transition between a valley region, where the urban area of Ciudad Victoria is located, and the foothills of the Sierra Madre Oriental.

- Cloud-top Height

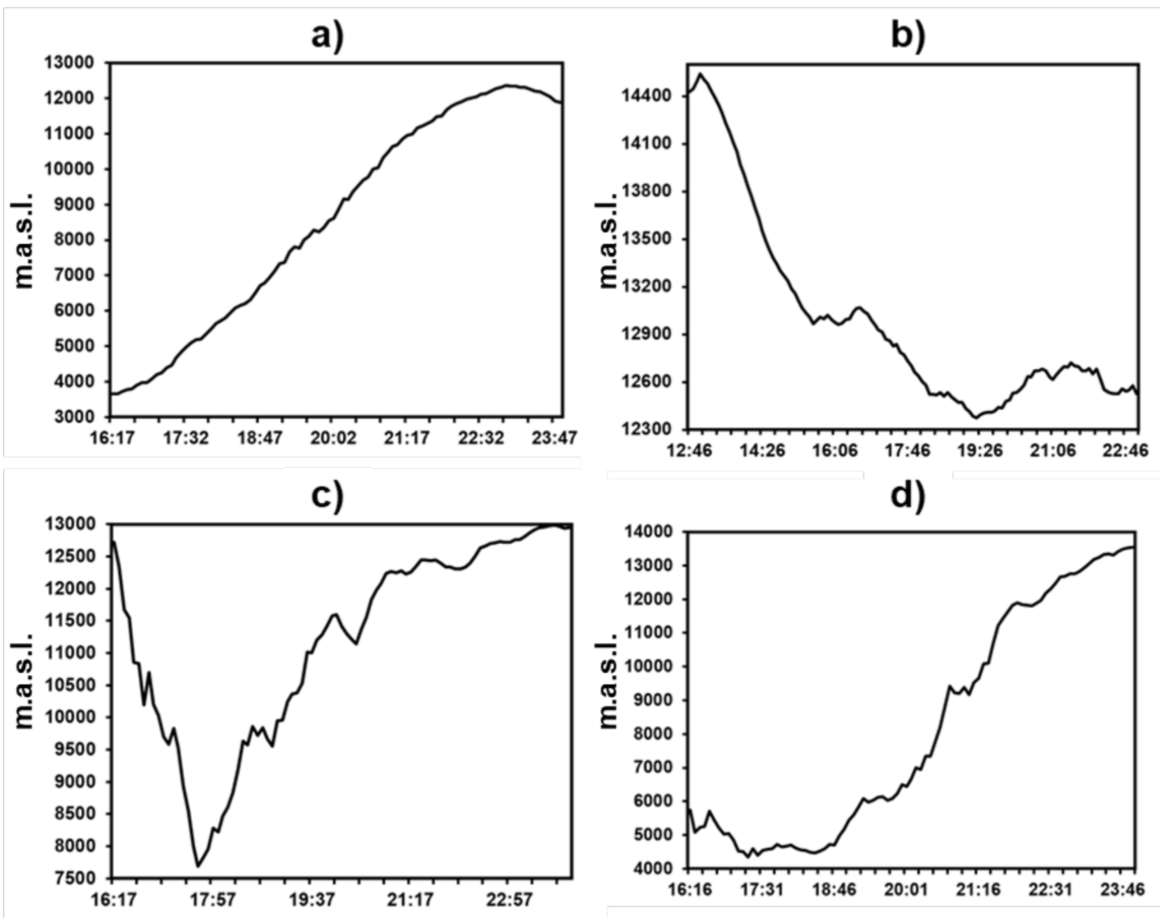


Figure 51. Temporal behavior of Cloud-top Height (CtH) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps).

CtH refers to the height or elevation at which the top of a cloud extends within the Earth's atmosphere. It is a critical parameter in meteorology and atmospheric science because it provides insight into clouds' vertical extent and structure. Specifically, higher cloud-top heights are directly related to heavier rainfall

(Snodgrass et al., 2009) and have also been linked to temperature and moisture layer heights, closely related to tropical cyclone intensity. CtH is also a key parameter in determining thunderstorm intensity through observations of growth rate (Castro et al., 2020).

Figure 51 presents the temporal evolution of CtH for Events 02, 05, 07, and 09. In Event 01 (Fig.51 a), a significant increase in the CtT values typical of the formation of a deep convective cloud is observed, which is characterized by a large vertical development (Figure 52). On the contrary, Event 05 (Fig. 51b) shows the decay effect of a mature cloud (or cloud cluster), with a gradual decrease in the CtH values. Event 07 (Figure51c) shows a minimum in CtH, followed by a sudden ascent, reaching its peak at 13,000 m. a.s.l. Meanwhile, in Event 09 (Figure51 d), a typical pattern of deep convective activity is observed, with consistent vertical growth.

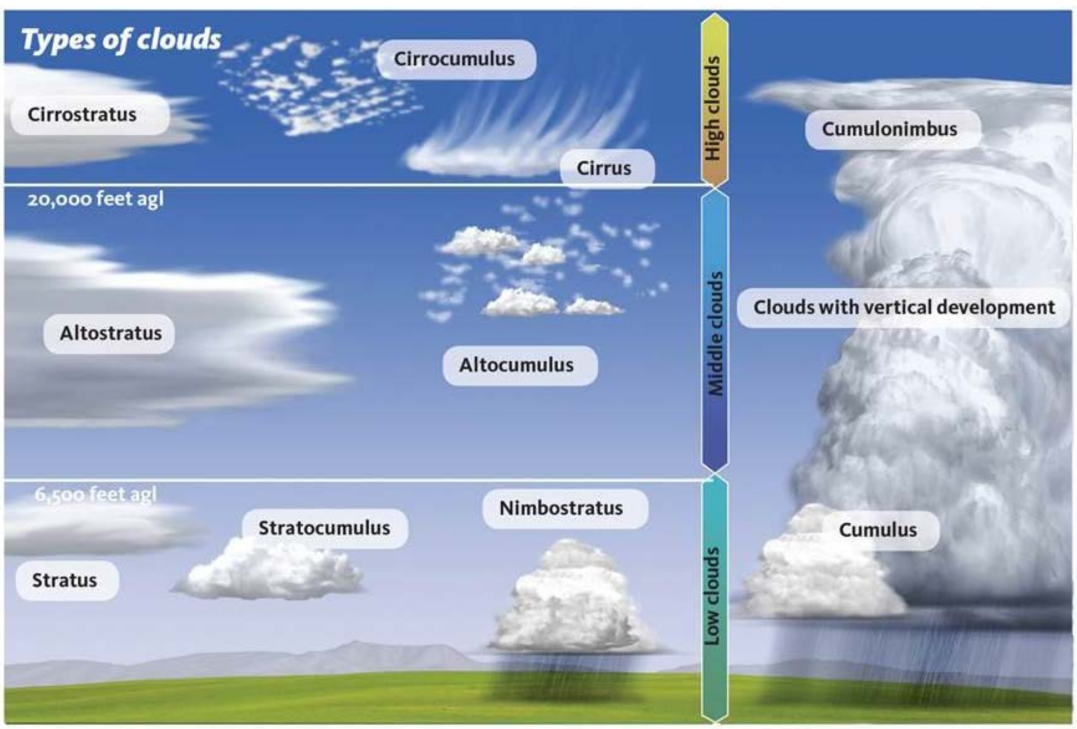


Figure 52. Schematization of type of clouds. Source: https://www.aopa.org/news-and-media/all-news/2016/august/flight-training/weather.

Figure 53 presents a 3D representation of the topography at the top of the cloud, recorded at the time of the highest mean CtH value. The cloud tops depicted in cool tones indicate the possible presence of Overshooting Cloud Top (OT) formations. OTs are dome-shaped protrusions that shoot from the top of a thunderstorm's anvil into the lower stratosphere (Shenk, 1974; Figure 54). These formations indicate intense convective activity within thunderstorms and are often associated with severe weather phenomena such as large hail, damaging winds, and tornadoes.

In Event 02 (Figure 53), three potential deep convective clouds are observed. These formations initially develop in the foothills of the mountain range, likely due to forced convection mechanisms. As the convective event progresses, these individual clouds merge into a unified system.

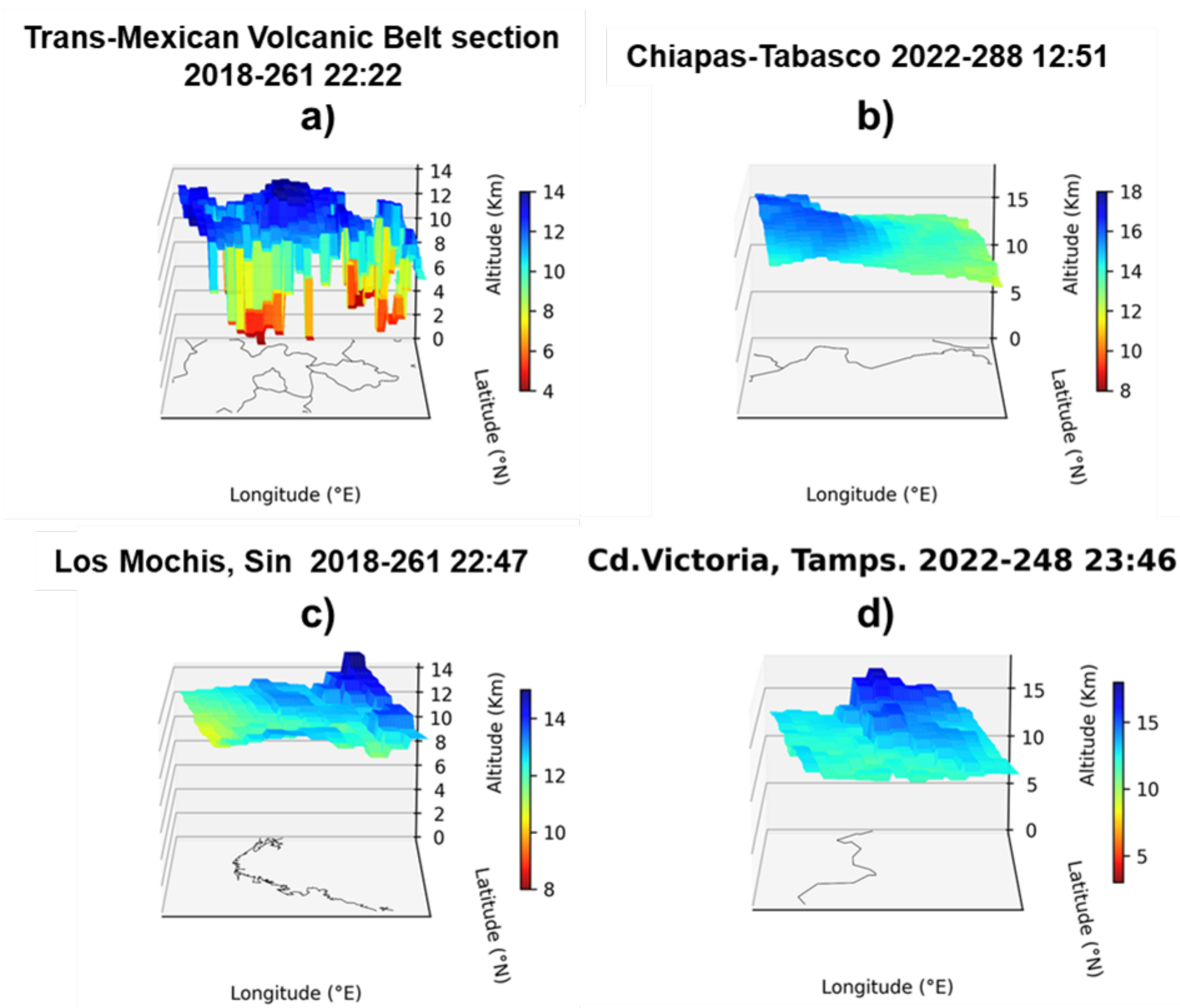


Figure 53. 3D Representation of the Cloud-top Height (CtH) for Event 02 (Trans-Mexican Volcanic Belt section; a), Event 05 (Chiapas-Tabasco; b), Event 07 (Los Mochis, Sin; c), and Event 09 (Ciudad Victoria, Tamps; d).

- Cloud-top Pressure

During the deep convective process, warm air near the surface becomes less dense and begins to rise (Figure 54). This rising air creates an updraft. If the air is sufficiently moist, it will continue to rise because of the buoyancy force. As the air rises, it cools adiabatically due to the decrease in atmospheric pressure. If the air contains enough moisture, it will reach its dew point, causing water vapor to condense into tiny water droplets or ice crystals. This process releases latent heat, further warming the air and increasing its buoyancy. The process is similar to other convection mechanisms, such as forced convection.

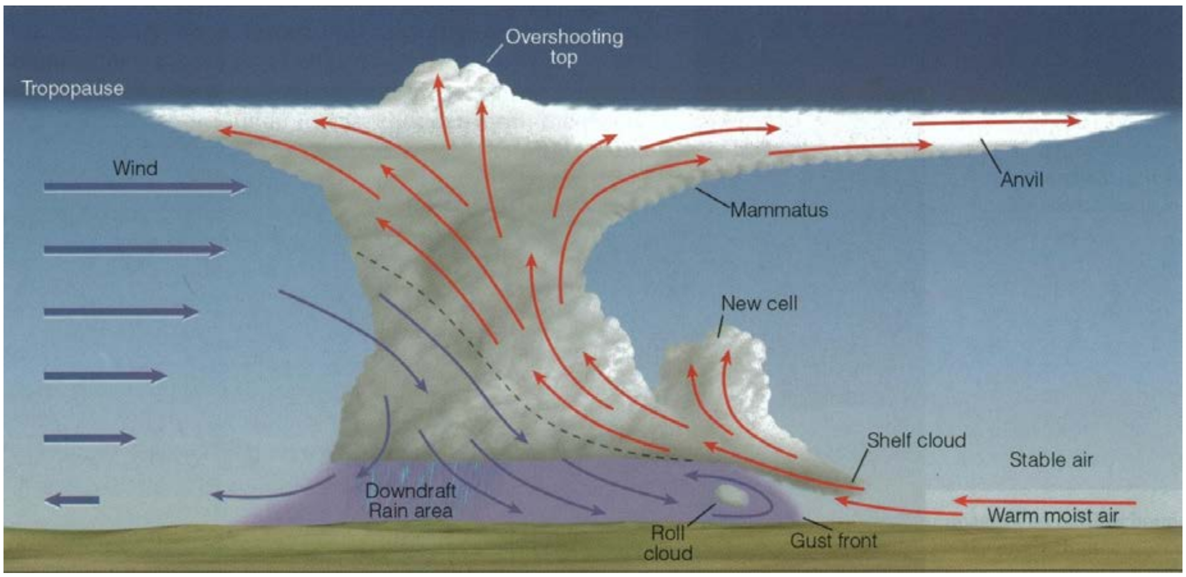


Figure 54. Diagram of a typical deep convective cloud. Source: https://weathertogether.net/weather-101/convective-clouds/.

The ascent of moist air initiates a suction effect, pulling in additional parcels of moist air while displacing dry air into the surrounding environment. Since high relative humidity air is lighter than dry air, this process causes a pressure drop. This pressure drop is particularly pronounced at the top of the cloud. Monitoring changes in cloud top pressure provides valuable insight into the intensity and potential hazards associated with convective weather systems, thereby helping to forecast severe weather events such as thunderstorms, hail, and tornadoes.

Figure 55 presents the temporal evolution of CtP for Events 02, 05, 07, and 09. A notable inverse relationship exists between the CtP and CtH values. This correlation results from deriving both variables using the Cloud Top Height algorithm for the ABI-GOES bands. (NOAA & NASA program, 2019). The ISCCP cloud classification (Rossow & Schiffer, 1991) designates a threshold of CtP<440 hPa for the identification of Cumulonimbus formations (Figure 11).

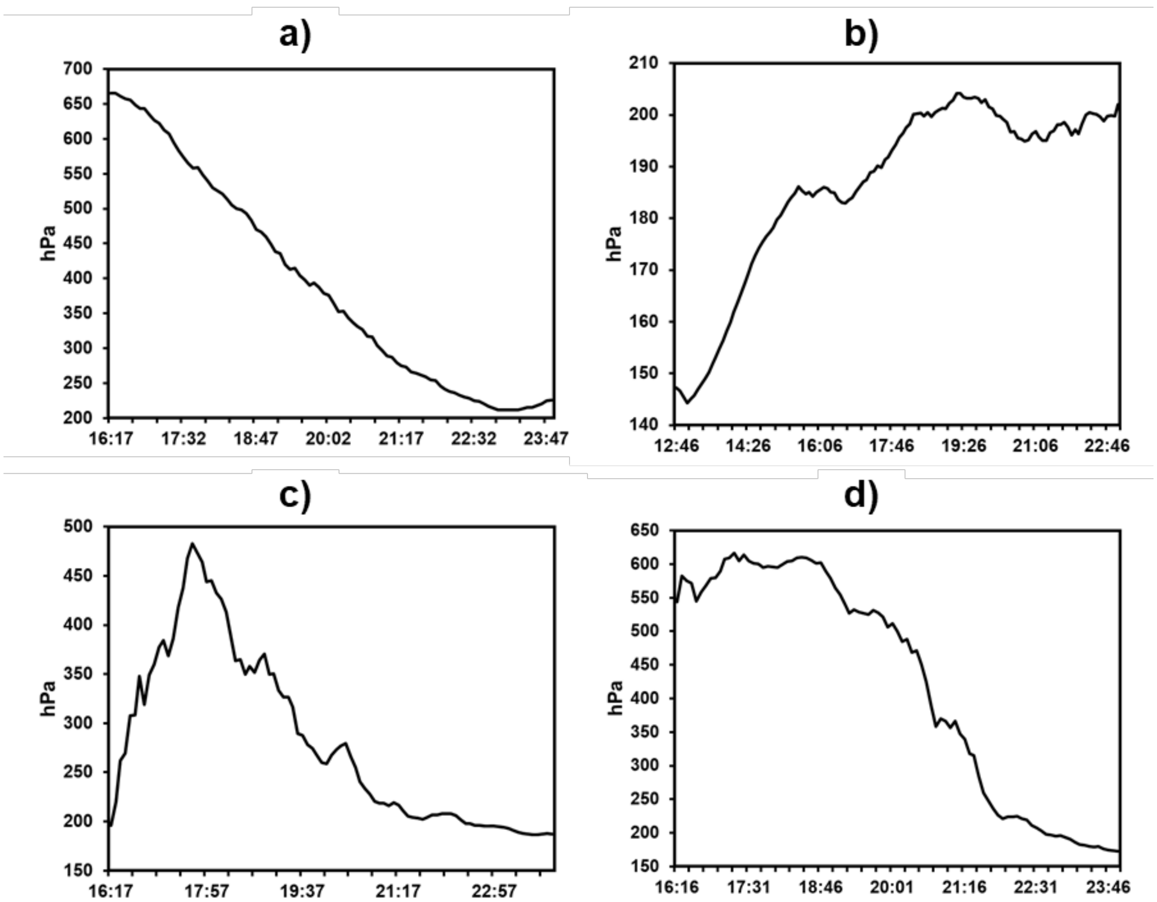


Figure 55. Temporal behavior of Cloud-top Pressure (CtP) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps).

- Cloud-top Temperature

Initially, as warm, moist air rises due to convection, it cools adiabatically with altitude, leading to the condensation of water vapor and the formation of cloud droplets. As the cloud continues to develop vertically, the CtT becomes a key indicator of the weather system's convective strength and potential severity. In the early stages of convective cloud formation, the cloud top temperature may be relatively warm compared to the surrounding environment. However, as the convective updrafts intensify and the cloud reaches higher altitudes, the CtT decreases due to the adiabatic cooling process (Dávila-Ortiz et al., 2024).

Figure 56 shows the temporal evolution of CtT for Events 02, 05, 07, and 09. Given that both the CtT and CtP products are derived from the CtH values generated with the ABI-GOES data, the trend of the mean temperature values follows the same pattern as the CtP values while being inversely proportional to the CtH values. Because of the adiabatic cooling effect, the coldest regions are directly related to the presence of vertically developing clouds (Figure 57).

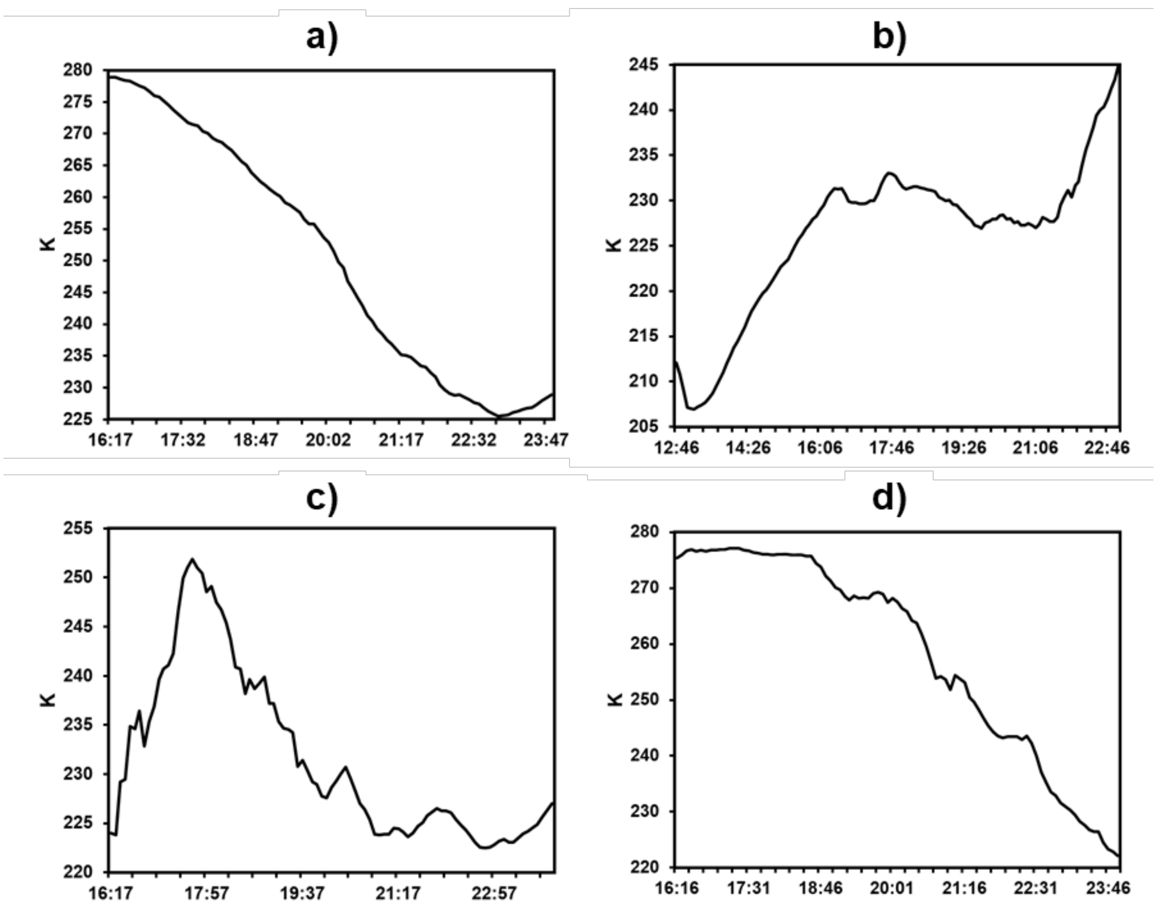


Figure 56. Temporal behavior of Cloud-top Temperature (CtT) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps).

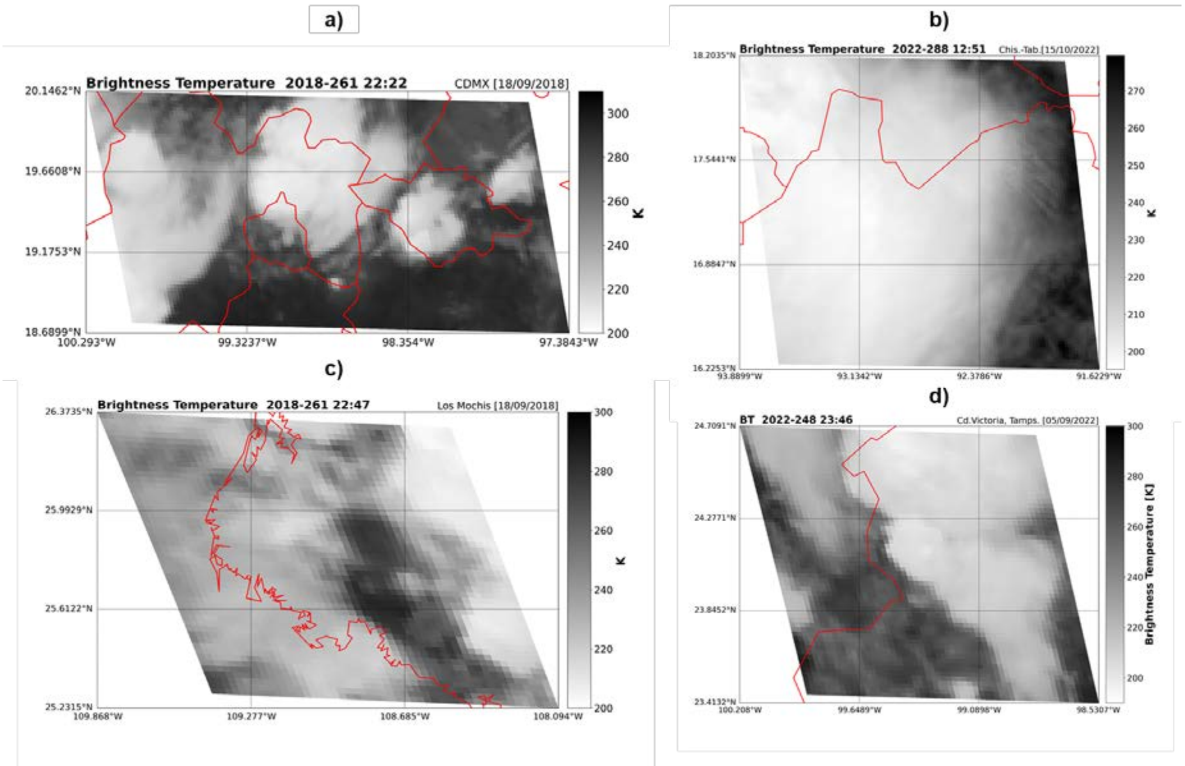


Figure 57. Cloud-top brightness temperature maps for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps).

In this study, it was estimated that storm clouds typically manifest when CtT values fall below a threshold of <220 K for the Los Mochis study site and <260 K for Mexico City (Figures 32 and 33). Notably, the temperature ranges observed in Events 02, 05, 07, and 09 predominantly fluctuated within these specified values. This finding underscores the consistency of the CtT thresholds as reliable indicators for identifying the onset of storm clouds at the respective locations.

Regarding the cloud-top brightness temperature map of Event 02 (Figure 57a), a remarkable correspondence is observed with the 3D representation of CtH (Figure 53a). The simultaneous development of four convective cores at 22:22 UTC can be observed in both cases. The regions with the highest elevation correspond to the coldest zones observed in the other analyzed events. The relief effect is particularly noticeable in Events 02 and 09, in which the generation of deep convective clouds by forced convection can be inferred.

It is imperative to highlight that the average values estimated during convective events for CtT, CtP, CtH, COD, and TPW include all zones within the analyzed spatial domain designated as CC (see Figure 49). This observation has significant implications. For instance, in Event 02, shown in Figure 57, a cloud system begins its development autonomously at the foothills of the Trans-Mexican Volcanic Belt section and subsequently evolves into a large-scale convective event. Also noteworthy is the entry and exit of cloud formations within the study domain due to the advection effect. Therefore, the plots of optical and cloud microphysical variables for this event show the study domain's general behavior instead of forming an independent event.

The study of independent deep convective clouds requires a smaller regionalization of the covered area or the use of conservative properties that serve as Lagrangian tracers to facilitate the tracking of isolated storm clouds.

- Cloud Optical Depth

The COD is a measure of the ability of a cloud to absorb and scatter incoming sunlight. Deep convective clouds often have high cloud optical depths due to their thickness and density. The presence of large water droplets, ice crystals, and other atmospheric particles within these clouds contributes to their high optical depth. As sunlight penetrates through the cloud, it interacts with these particles, leading to scattering and absorption of light.

The optical depth of deep convective clouds is influenced by several factors, including cloud thickness, cloud water content, particle size distribution, and vertical extent (American Meteorological Society, 2024). Optical depth is typically higher in mature convective clouds where the concentration of particles is greater and the cloud is thicker.

Figure 58 shows the temporal evolution of COD for Events 02, 05, 07, and 09. In the first instance, it can be seen that all events exceeding the threshold proposed by the ISCCP cloud classification (Rossow & Schiffer, 1991) of COT>23 are associated with larger vertical development clouds (considering the ABI-GOES derived product COD as equivalent to the MODIS COT product). Similar to the previous variables, these values represent averages over the entire simulation domain. Events 02, 07, and 09 show the typical behavior of the formation process of a deep convective cloud. In particular, the thickness of the cloud increases as it grows vertically (Figure 51), accompanied by an increase in moisture content, which drives a decrease in pressure values (Figure 55). In addition, adiabatic cooling leads to cooling of the cloud top (Figure 56).

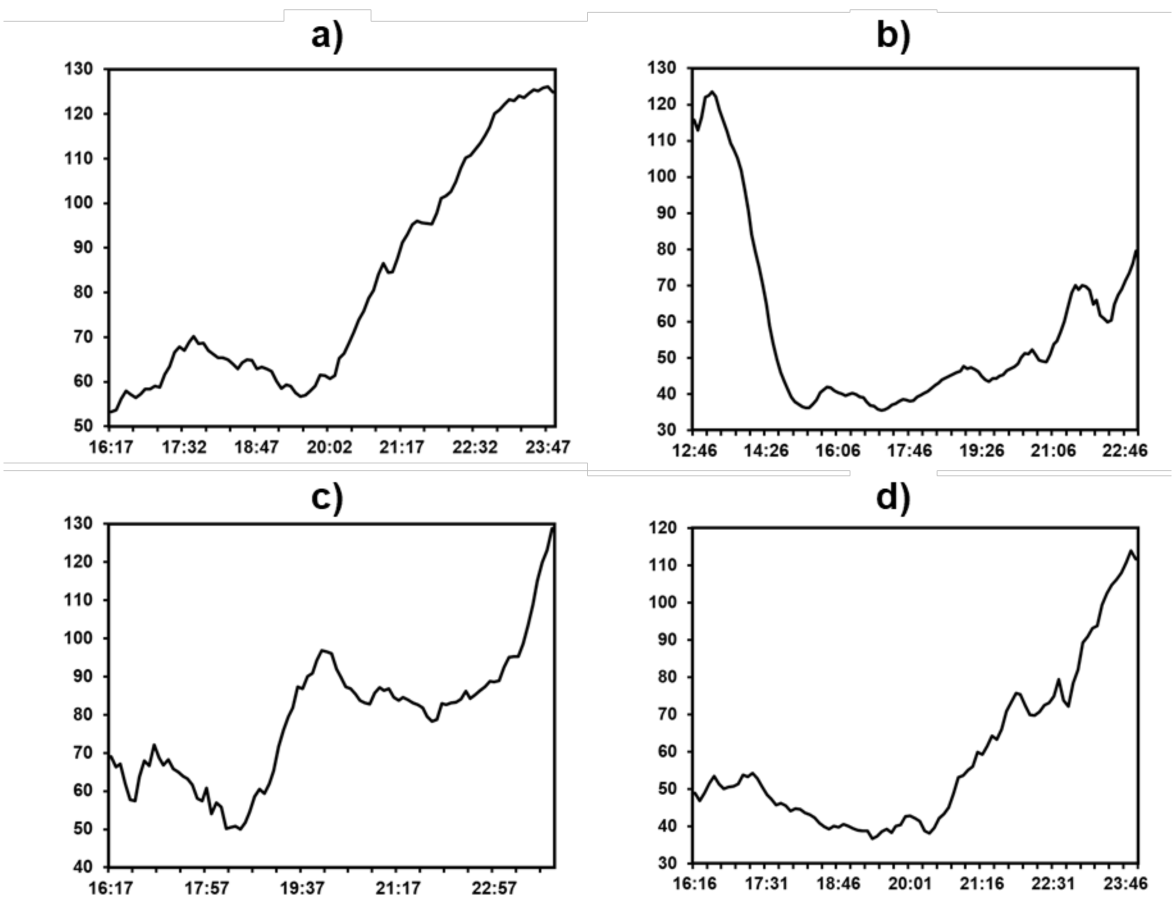


Figure 58. Temporal behavior of Cloud Optical Depth (COD) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps).

On the other hand, the COD values of Event 05 (Figure 58b), potentially indicative of the dissipation phase of a mature cloud, present an abrupt decrease in its depth values from COD greater than 120 to COD ≈ 40 by 15:00 UTC, consistent with the estimations of CtH, CtP, and CtT.

Figure 59 illustrates the average COD maps estimated over the entire duration of the convective event. In the case of Event 02 (Fig. 59 a), this map provides insight into the spatial distribution of the convective cores, revealing the presence of three distinct cores closely related to the areas of highest topographic elevation (Figure 50). In these cores, average values between 40 and 110 were estimated throughout the event.

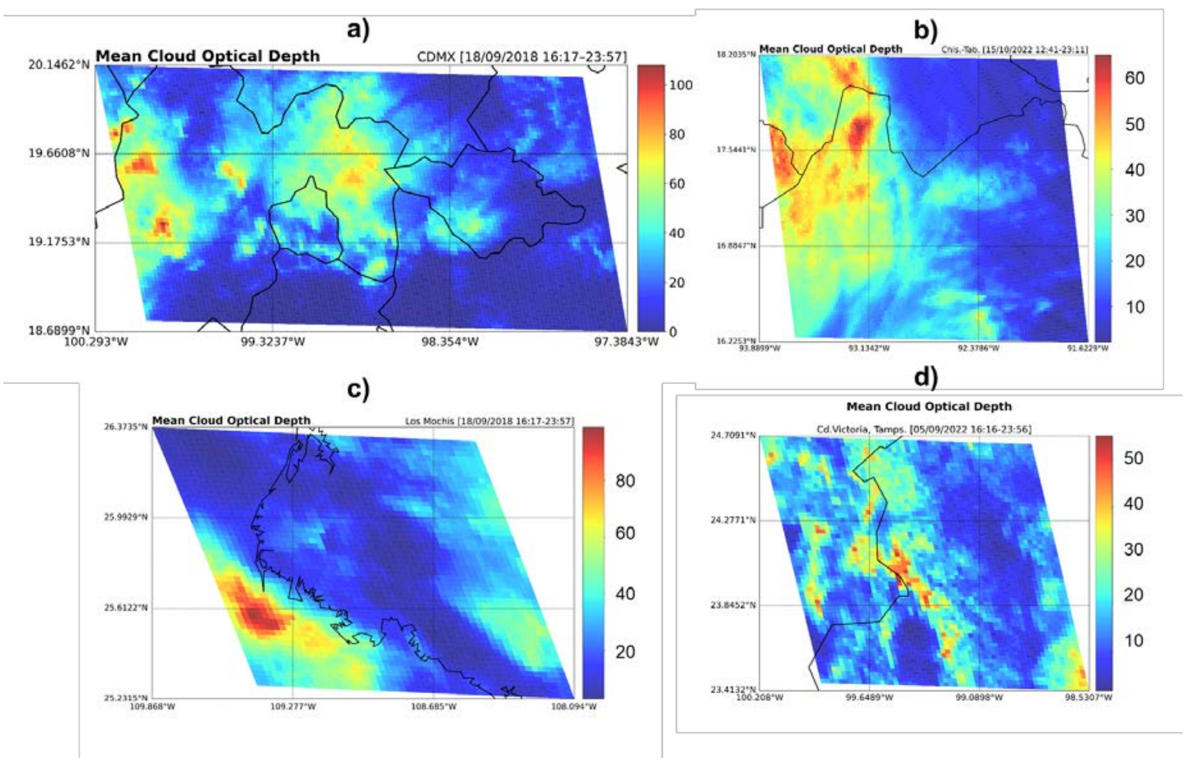


Figure 59. Mean Cloud Optical Depth (COD) maps for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps).

In contrast, in Event 09 (Fig. 59 d), the highest mean COD values were concentrated over the region with the highest topographic elevation, suggesting the influence of forced convection. Conversely, for Event 05 (Figure59 b), the peak COD values correspond to the eastern region of the simulation domain, indicating the location where the deep convective cloud before its dissipation phase.

- Total Precipitable Water

TPW represents the total amount of water vapor present in a vertical column of the atmosphere if the water vapor in that column condensed and precipitated. TPW provides valuable insight into the moisture content available for cloud development and precipitation in the context of deep convective clouds. The TPW product also initializes the moisture field used in numerical weather prediction models (NOAA & NASA program, 2019).

In this study, TPW emerges as a parameter of critical importance because its integration with historical rainfall data in this framework generates quasi-real-time precipitation scenarios. In this sense, Figure 60 shows the estimated TPW values for each analyzed convective event.

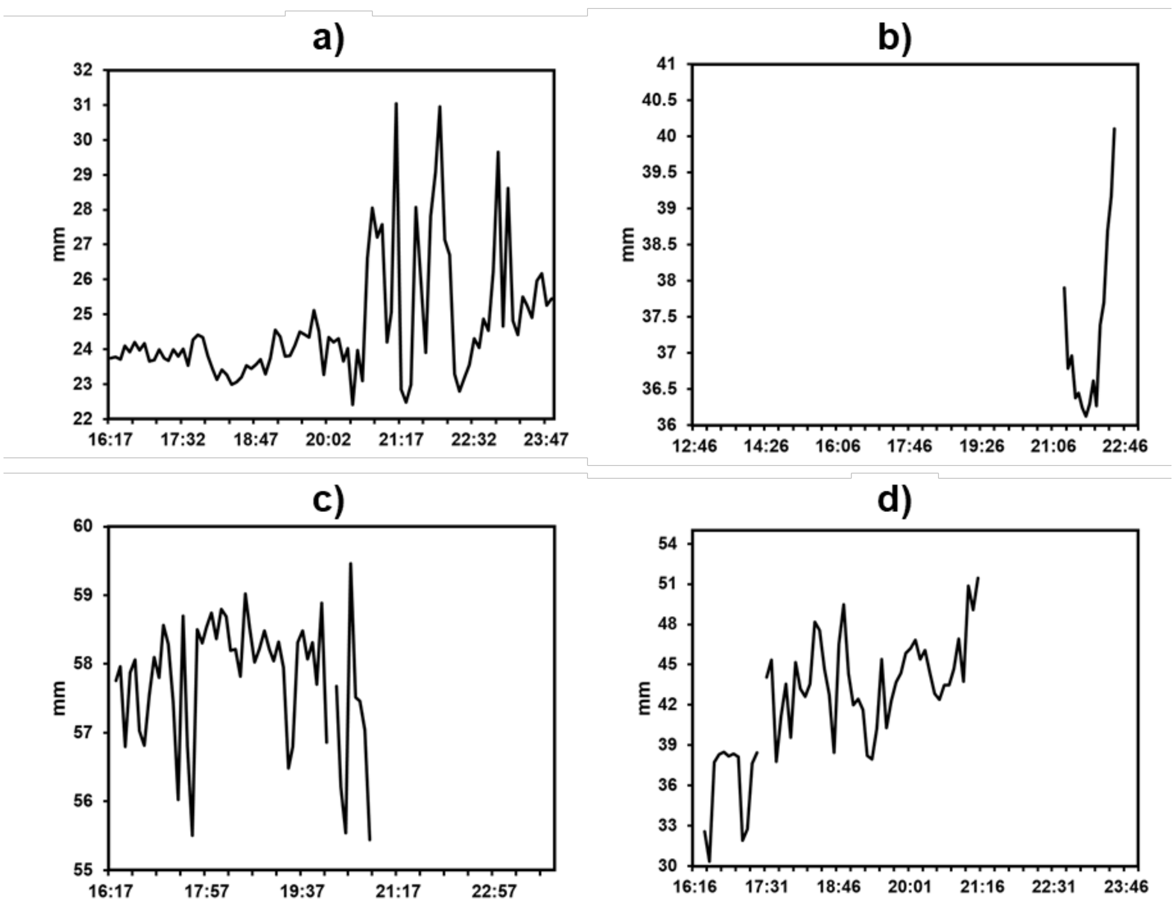


Figure 60. Temporal behavior of Total Precipitable Water (TPW) for the Event 02 (a; Trans-Mexican Volcanic Belt section), Event 05 (b; Chiapas-Tabasco), Event 07 (c; Los Mochis, Sin), and Event 09 (d; Ciudad Victoria, Tamps).

First, Event 02 (Figure 60a) shows TPW values between 22 and 32 mm, with three important peaks occurring after 20:00 UTC, coinciding with intensified growth rates of COD values (Figure 59a). On the other hand, for most of Event 05 (Figure 60c), TPW data are not detected, but in its final stages, a maximum of 40 mm is recorded. In this case, the CtH, CtP, and CtT estimates do not show the typical behavior of a deep convection event, although the COD values are still above 50.

Event 07 (Figure 60c), associated with Tropical Depression 19-E event (Dávila Ortiz, 2019), showed significant differences in TPW values compared with those estimated from the ML model-based framework (Figure 45a). However, both simulated events agreed on the order of magnitude, oscillating between 55 and 60 mm, and both simulated events approximately stopped detecting TPW values after 21:00 UTC, although the COD peak coincided with the stage of maximum storm cloud maturity (Figure 58). Similarly, Event 09 (Figure60d) follows this pattern, with no TPW values recorded during the peak growth phase of the convective event.

6.3. Evolution of vertical development clouds (Velocity, Acceleration and Energy)

Buoyancy is the upward force exerted on an object immersed in a fluid (liquid or gas) because of the difference in density between the object and the fluid. In the context of deep convection, buoyancy plays an important role in driving the upward motion of air parcels within convective clouds. As air near the surface is heated, it becomes less dense and therefore more buoyant than the surrounding cooler air. This buoyant air parcel rises due to the buoyancy force exerted by the denser surrounding air.

As the buoyant air parcel rises, it expands and cools as a result of the decrease in atmospheric pressure with height. If the temperature of the parcel remains higher than that of its surroundings, it will continue to become less dense and therefore continue to rise. This process of rising and cooling can influence the formation of deep convective clouds.

Within deep convective clouds, buoyancy continues to drive the upward motion of the air parcels. As the moist air rises, it cools and condenses, releasing the latent heat of condensation. This latent heat further reduces the density of the air parcel, increasing its buoyancy and promoting even greater upward motion. This cycle of rising, condensing, and releasing latent heat sustains the deep convection process, leading to the development of towering cumulonimbus clouds and potentially severe weather phenomena such as thunderstorms, heavy rainfall, and even tornadoes.

For these reasons, the analysis of this property along with parameters such as CtH, CtP, CtT, COD, and TPW allows meteorologists and forecasters to gain deeper insights into the underlying mechanisms driving convective hazards, encompassing hurricanes and other mesoscale events.

In this proposed framework, there is a focus on estimating buoyancy within pixels classified as CC class. This estimation involves calculating the rate and acceleration of convection derived from the CtH values' rate of change over 5-minute intervals. The mathematical relationship between the rate and acceleration of convection and Buoyancy Force is outlined below.

Considering Newton's 2nd Law:

$$\vec{\mathbf{F}} = m\vec{\mathbf{a}} \qquad (Eq. 21)$$

$$\vec{\mathbf{F}} = \vec{\mathbf{F}}_1 + \vec{\mathbf{F}}_1 + \cdots + \vec{\mathbf{F}}_n \qquad (Eq. 22)$$

Vector $\vec{\mathbf{F}}$ is the resultant of the interaction of several forces. In this mathematical derivation, several of these forces are neglected, such as the vertical pressure gradient , $\left(-\frac{1}{\rho}\frac{\partial p}{\partial z}\right)$ assuming equilibrium conditions.

$$\vec{\mathbf{a}} = \frac{\mathrm{d}\vec{\mathbf{v}}}{\mathrm{d}t} \qquad (Eq. 23)$$

Substituting Eq. 23 into Eq. 21.

$$\vec{\mathbf{F}} = m\frac{\mathrm{d}\vec{\mathbf{v}}}{\mathrm{d}t} \qquad (Eq. 24)$$

The velocity vector is composed of ***x***, ***y***, and ***z*** components.

$$\vec{\mathbf{v}} = \hat{\imath}u + \hat{\jmath}v + \hat{k}w \qquad (Eq. 25)$$

The convection process occurs due to the ascend of air parcels, so only the effect of the vertical velocity is analyzed, which is as follows:

$$\mathbf{v}_z = w = \frac{\mathrm{d}z}{\mathrm{d}t} \qquad (Eq. 26)$$

Substituting Eq. 26 into Eq. 24.

$$\vec{\mathbf{F}} = m\frac{\mathrm{d}w}{\mathrm{d}t} \qquad (Eq. 27)$$

$$\frac{\vec{\mathbf{F}}}{m} = \frac{\mathrm{d}w}{\mathrm{d}t} \qquad (Eq. 28)$$

The term force per unit mass (assuming m=1kg) represents the Buoyancy Force (*B*).

$$B = \frac{\mathrm{d}w}{\mathrm{d}t} = a_z \qquad (Eq. 29)$$

Therefore, the effect of the Buoyancy Force can be observed in the differential of the vertical velocity vector or the acceleration in the z direction.

The buoyancy force is particularly evident in the behavior of the Rate and Acceleration of Convection estimates (Figure 61). Across all events, an oscillatory pattern is observed that is typical of the complex nature of the cloud formation system detected within each simulation domain and the effect of turbulence and atmospheric instability.

In this context, turbulence can facilitate the vertical rise of air, while at the same time promoting mixing and dispersion, which affects the availability of water vapor for cloud formation. However, turbulence can also contribute to cloud dispersal, particularly in the upper layers of the atmosphere, where it can break clouds into smaller fragments or disperse them completely.

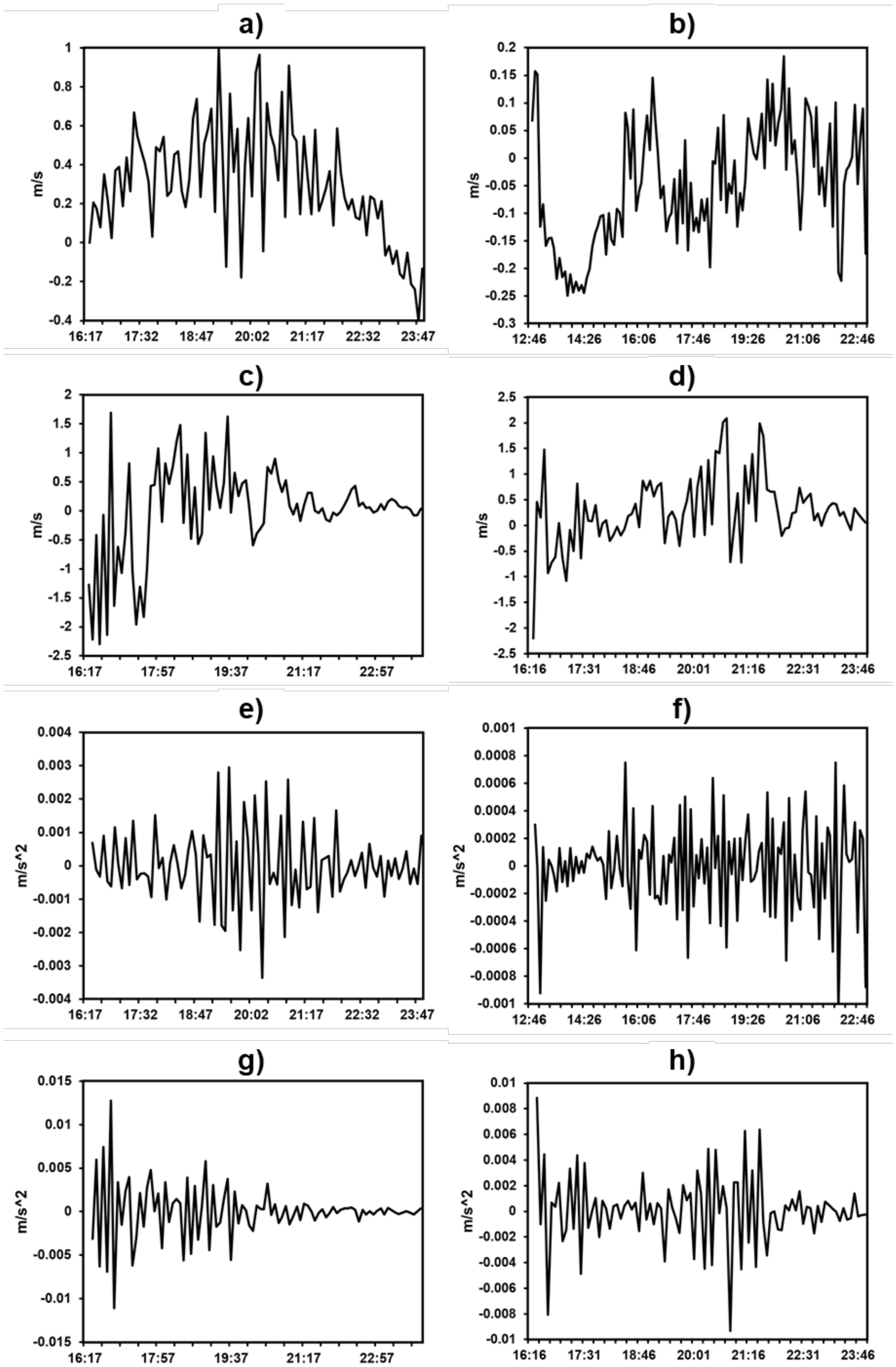


Figure 61. Rate and Acceleration of Convection for the Event 02 (a, e; Trans-Mexican Volcanic Belt section), Event 05 (b, f; Chiapas-Tabasco), Event 07 (c, g; Los Mochis, Sin), and Event 09 (d, h; Ciudad Victoria, Tamps).

Chapter 7. Conclusions

This study developed a quasi-real-time framework for detecting and monitoring deep convective events based on ABI-GOES data. This framework was guided by three key principles: using open-source software and open-access products, ensuring automation, and scalability. Operating cyclically and autonomously at 5-minute intervals and aligning the temporal resolution of the ABI bands, this modeling framework represents a significant advance in deep convective cloud monitoring. This proposal is novel as it addresses the use of ABI-GOES data to generate real-time risk scenarios for storms and flooding, a method not covered in any previous work. Specifically, in the case of Mexico, this type of data is not yet utilized for real-time forecasting tasks, highlighting a significant opportunity for improvement in the region's weather prediction capabilities.

This framework is based on the use of ML techniques with eight different models and ensemble strategies, including LR, RF, MLP, LRstacking, RFstacking, Bagging, Hard-Voting, and Soft-Voting. These models were evaluated at two study sites: Los Mochis and Mexico City. These sites were selected for their intense convective activity and high vulnerability to extreme weather events. The results for the testing dataset indicate that relatively simple approaches such as LR or RF show promising performance for the identification and simulation of deep convective clouds in both study areas. achieving POD ≈ 0.84 for Los Mochis and POD ≈ 0.72 for Mexico City. Simultaneously, FAR ≈ 0.2 and 0.4 were estimated, respectively.

Furthermore, the efficacy of implementing a post-processing filter based on the lightning incidence recorded by GLM-GOES was evaluated in order to enhance the detection capabilities of the ML models. In this context, although the impact did not represent a significant change for Los Mochis, a significant improvement was observed for all ML approaches in Mexico City. This can be attributed to several factors, including the complexity of the terrain and its atmospheric dynamics, the potential impact of the proportion of CC and noCC classes on the overall performance of the models, and limitations in labeling reference events from MODIS data.

Challenges related to the simulation of cloud edges were identified and attributed to the limitations of pixel-based classification methods. Nevertheless, the overall performance of most ML approaches remained commendable. Future work includes implementing models that use contextual spatial information, such as CNNs, and exploring unsupervised learning techniques, such as storm cloud clustering, as an alternative to constructing labeled datasets.

Moreover, this research integrates a deep convective cloud detection framework with multi-source precipitation data by generating dimensionless curves from historical records at the Los Mochis study site and estimating the mean Total Precipitable Water in deep convective clouds. This integration produces precipitation scenarios applicable to urban flooding and river overflow analysis, hydraulic design, and landslide risk assessment, among other applications.

The methodological framework proposed in this research, which is scalable and can be applied to cover all of Mexico. It serves as a valuable tool for risk managers, decision-makers, and vulnerable population sectors. It has been applied to the study and analysis of storm cloud formation processes and their physical properties in different regions of Mexico.

This work establishes a precedent for the implementation of an Early Warning System (EWS) for hazards associated with intense convective activity in Mexico, a region characterized by complex atmospheric dynamics and high vulnerability to extreme hydrometeorological conditions. It represents a pioneering effort to bridge the gap created by the limited severe weather monitoring infrastructure and the highly heterogeneous distribution of ground-based sensors.

Although this framework has focused on the identification of deep convective clouds, the estimation of precipitation scenarios derived from these formations, and the zoning of potential flood hazard areas, the system has the potential to be implemented in a wide range of risk assessments, including those related to lightning, hail, tornado generation, and even cloud seeding applications.

In summary, this dissertation contributes to remote sensing and machine learning of meteorological parameters by providing innovative approaches to deep convective cloud monitoring and advancing our understanding of convective hazards.

References

Adrianto, I., Trafalis, T. B., & Lakshmanan, V. (2009). Support vector machines for spatiotemporal tornado prediction. *International Journal of General Systems*, *38*(7), 759–776. https://doi.org/10.1080/03081070601068629

Afzali Gorooh, V., Kalia, S., Nguyen, P., Hsu, K., Sorooshian, S., Ganguly, S., & Nemani, R. R. (2020). Deep Neural Network Cloud-Type Classification (DeepCTC) Model and Its Application in Evaluating PERSIANN-CCS. *Remote Sensing*, *12*(2), Article 2. https://doi.org/10.3390/rs12020316

Agbehadji, I. E., Mabhaudhi, T., Botai, J., & Masinde, M. (2023). A Systematic Review of Existing Early Warning Systems' Challenges and Opportunities in Cloud Computing Early Warning Systems. *Climate*, *11*(9), Article 9. https://doi.org/10.3390/cli11090188

Ahijevych, D., Pinto, J. O., Williams, J. K., & Steiner, M. (2016). Probabilistic Forecasts of Mesoscale Convective System Initiation Using the Random Forest Data Mining Technique. *Weather and Forecasting*, *31*(2), 581–599. https://doi.org/10.1175/WAF-D-15-0113.1

Akbari Asanjan, A., Yang, T., Hsu, K., Sorooshian, S., Lin, J., & Peng, Q. (2018). Short-Term Precipitation Forecast Based on the PERSIANN System and LSTM Recurrent Neural Networks. *Journal of Geophysical Research: Atmospheres*, *123*(22), 12,543-12,563. https://doi.org/10.1029/2018JD028375

American Meteorological Society. (2023). *Overshooting top—Glossary of Meteorology*. Glossary of Meteorology. https://glossary.ametsoc.org/wiki/Overshooting_top

American Meteorological Society. (2024). *cloud optical depth—Glossary of Meteorology*. Glossary of Meteorology. https://glossary.ametsoc.org/wiki/Cloud_optical_depth#:~:text=The%20vertical%20optical%20thickness%20between,blackbodies%20in%20the%20thermal%20infrared.

Anderson, C. J., & Arritt, R. W. (1998). Mesoscale Convective Complexes and Persistent Elongated Convective Systems over the United States during 1992 and 1993. *Monthly Weather Review*, *126*(3), 578–599. https://doi.org/10.1175/1520-0493(1998)126<0578:MCCAPE>2.0.CO;2

Barrett, B. S., & Gensini, V. A. (2013). Variability of central United States April–May tornado day likelihood by phase of the Madden-Julian Oscillation. *Geophysical Research Letters*, *40*(11), 2790–2795. https://doi.org/10.1002/grl.50522

Barrett, B. S., & Henley, B. N. (2015). Intraseasonal Variability of Hail in the Contiguous United States: Relationship to the Madden–Julian Oscillation. *Monthly Weather Review*, *143*(4), 1086–1103. https://doi.org/10.1175/MWR-D-14-00257.1

Bartos, M., Wong, B., & Kerkez, B. (2018). Open storm: A complete framework for sensing and control of urban watersheds. *Environmental Science: Water*

Research & Technology, *4*(3), 346–358. https://doi.org/10.1039/C7EW00374A
Bedka, K., Brunner, J., Dworak, R., Feltz, W., Otkin, J., & Greenwald, T. (2010). Objective Satellite-Based Detection of Overshooting Tops Using Infrared Window Channel Brightness Temperature Gradients. *Journal of Applied Meteorology and Climatology*, *49*(2), 181–202. https://doi.org/10.1175/2009JAMC2286.1
Bedka, K. M., Allen, J. T., Punge, H. J., Kunz, M., & Simanovic, D. (2018). A Long-Term Overshooting Convective Cloud-Top Detection Database over Australia Derived from MTSAT Japanese Advanced Meteorological Imager Observations. *Journal of Applied Meteorology and Climatology*, *57*(4), 937–951. https://doi.org/10.1175/JAMC-D-17-0056.1
Bedka, K. M., & Khlopenkov, K. (2016). A Probabilistic Multispectral Pattern Recognition Method for Detection of Overshooting Cloud Tops Using Passive Satellite Imager Observations. *Journal of Applied Meteorology and Climatology*, *55*(9), 1983–2005. https://doi.org/10.1175/JAMC-D-15-0249.1
Bonta, J. (2004). Development and utility of Huff curves for disaggregating precipitation amounts. *Applied Engineering in Agriculture*, *20*. https://doi.org/10.13031/2013.17467
Breiman, L. (1984). *Classification and Regression Trees* (1st Edition). Routledge. https://doi.org/10.1201/9781315139470
Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655
Breiman, L. (1999). Pasting Small Votes for Classification in Large Databases and On-Line. *Machine Learning*, *36*(1), 85–103. https://doi.org/10.1023/A:1007563306331
Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
Brooks, H. E. (2013). Severe thunderstorms and climate change. *Atmospheric Research*, *123*, 129–138. https://doi.org/10.1016/j.atmosres.2012.04.002
Burke, A., Snook, N., Ii, D. J. G., McCorkle, S., & McGovern, A. (2020). Calibration of Machine Learning–Based Probabilistic Hail Predictions for Operational Forecasting. *Weather and Forecasting*, *35*(1), 149–168. https://doi.org/10.1175/WAF-D-19-0105.1
Calvello, M., & Piciullo, L. (2016). Assessing the performance of regional landslide early warning models: The EDuMaP method. *Natural Hazards and Earth System Sciences*, *16*(1), 103–122. https://doi.org/10.5194/nhess-16-103-2016
Cancelada, M., Salio, P., Vila, D., Nesbitt, S. W., & Vidal, L. (2020). Backward Adaptive Brightness Temperature Threshold Technique (BAB3T): A Methodology to Determine Extreme Convective Initiation Regions Using Satellite Infrared Imagery. *Remote Sensing*, *12*(2), Article 2. https://doi.org/10.3390/rs12020337
Carbajal, N., León-Cruz, J. F., Pineda-Martínez, L. F., Tuxpan-Vargas, J., & Gaviño-Rodríguez, J. H. (2019). Occurrence of Anticyclonic Tornadoes in a Topographically Complex Region of Mexico. *Advances in Meteorology*, *2019*, e2763153. https://doi.org/10.1155/2019/2763153

Castro, E., Ishida, T., Takahashi, Y., Kubota, H., Perez, G. J., & Marciano, J. S. (2020). Determination of Cloud-top Height through Three-dimensional Cloud Reconstruction using DIWATA-1 Data. *Scientific Reports*, *10*(1), 7570. https://doi.org/10.1038/s41598-020-64274-z

Cavazos, T., Turrent, C., & Lettenmaier, D. P. (2008). Extreme precipitation trends associated with tropical cyclones in the core of the North American monsoon. *Geophysical Research Letters*, *35*(21). https://doi.org/10.1029/2008GL035832

Centro Nacional de Prevención de Desastres. (2023, June 16). *Servicio integral de envío masivo de mensajes CAP* [A distancia, transmisión en vivo]. Convocatoria 2023. Sistemas de Alerta Temprana. https://www.gob.mx/cenapred/documentos/convocatoria-2023-sistemas-de-alerta-temprana

Chen, H., Chandrasekar, V., Cifelli, R., & Xie, P. (2020). A Machine Learning System for Precipitation Estimation Using Satellite and Ground Radar Network Observations. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(2), 982–994. https://doi.org/10.1109/TGRS.2019.2942280

Chow, V. T. (1959). *Open-channel hydraulics* (Vol. 1). McGraw-Hill. http://science.sciencemag.org/content/131/3408/1215.2

Chow, V. T., Maidment, D. R., Mays, L. W., Saldarriaga, J. G., & others. (1994). *Hidrología aplicada*. McGraw Hill.

CHRS. (2023). *CHRS Data Portal* [Official Website]. CHRS Data Portal. https://chrsdata.eng.uci.edu/

Cintineo, J. L., Pavolonis, M. J., Sieglaff, J. M., Wimmers, A., Brunner, J., & Bellon, W. (2020). A Deep-Learning Model for Automated Detection of Intense Midlatitude Convection Using Geostationary Satellite Images. *Weather and Forecasting*, *35*(6), 2567–2588. https://doi.org/10.1175/WAF-D-20-0028.1

CLICOM. (2016). *Base de Datos Climatológica Nacional. Servicio Meteorológico Nacional a través del CICESE*. [Base de datos]. http://clicom-mex.cicese.mx.

Conde, A. C., Hernández, G. E., Esquivel, N., López, J., López, F., Montes, R., Nava, Y., & Ruíz, K. (2016). Variabilidad climática y escenarios de cambio climático. Herramientas para los estudios de impactos potenciales y vulnerabilidad actual y futura. Ejemplos para México, Centroamérica y El Caribe. In A. C. Conde & J. López (Eds.), *Variabilidad y Cambio Climático. Impactos, Vulnerabilidad y Adaptación al Cambio Climático en América Latina y el Caribe* (pp. 25–72). INE. http://biblioteca.semarnat.gob.mx/janium/Documentos/Ciga/Libros2013/CD002498.pdf#page=27

Cools, J., Vanderkimpen, P., El Afandi, G., Abdelkhalek, A., Fockedey, S., El Sammany, M., Abdallah, G., El Bihery, M., Bauwens, W., & Huygens, M. (2012). An early warning system for flash floods in hyper-arid Egypt. *Natural Hazards and Earth System Sciences*, *12*(2), 443–457. https://doi.org/10.5194/nhess-12-443-2012

da Silva Neto, C. P., Alves Barbosa, H., & Assis Beneti, C. A. (2016). A method for convective storm detection using satellite data. *Atmósfera*, *29*(4), 343–358. https://doi.org/10.20937/ATM.2016.29.04.05

Dávila Ortiz, R. (2019). *Respuesta hidrológica en zonas urbanas ante eventos severos de tormenta aplicando modelación numérica e información SAR: Caso Los Mochis* [masterThesis, Instituto Potosino de Investigación Científica y Tecnológica, A.C.]. https://repositorio.ipicyt.edu.mx///handle/11627/5259

Dávila Ortiz, R., Tuxpan Vargas, J., & Velázquez Zapata, J. A. (2023a). Identification of Deep Convection Clouds Using ABIGOES Data and Machine Learning Techniques: The Case of Los Mochis, Sinaloa, Mexico. *2023 Mexican International Conference on Computer Science (ENC)*, 1–7. https://doi.org/10.1109/ENC60556.2023.10508667

Dávila Ortiz, R., Tuxpan Vargas, J., & Velázquez Zapata, J. A. (2023b). La producción del riesgo ante eventos de inundación en Los Mochis, Sinaloa. In *Riesgos y desastres relacionados con agua transformación del territorio, inundaciones y contaminación.* El Colegio de San Luis, A.C.

Dávila-Ortiz, R., Carbajal-Pérez, J. N., Velázquez-Zapata, J. A., & Tuxpan-Vargas, J. (2024). Approximation of a Convective-Event-Monitoring System Using GOES-R Data and Ensemble ML Models. *Remote Sensing*, *16*(4), Article 4. https://doi.org/10.3390/rs16040675

Dixon, M., & Wiener, G. (1993). TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-based Methodology. *Journal of Atmospheric and Oceanic Technology*, *10*(6), 785–797. https://doi.org/10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2

Eakin, H., Lerner, A. M., Manuel-Navarrete, D., Hernández Aguilar, B., Martínez-Canedo, A., Tellman, B., Charli-Joseph, L., Fernández Álvarez, R., & Bojórquez-Tapia, L. (2016). Adapting to risk and perpetuating poverty: Household's strategies for managing flood risk and water scarcity in Mexico City. *Environmental Science & Policy*, *66*, 324–333. https://doi.org/10.1016/j.envsci.2016.06.006

El Debate. (2018, September 20). Los Mochis, Sinaloa bajo el agua, estos son los estragos de las fuertes lluvias. *El Debate*. https://www.debate.com.mx/losmochis/lluvias-inundaciones-sinaloa-los-mochis-el-fuerte-clima-mexico-20180920-0011.html

El- Sayed, E. A. H. (2018). Development of synthetic rainfall distribution curves for Sinai area. *Ain Shams Engineering Journal*, *9*(4), 1949–1957. https://doi.org/10.1016/j.asej.2017.01.010

Elfeki, A. M., Ewea, H. A., & Al-Amri, N. S. (2014). Development of storm hyetographs for flood forecasting in the Kingdom of Saudi Arabia. *Arabian Journal of Geosciences*, *7*(10), 4387–4398. https://doi.org/10.1007/s12517-013-1102-3

Forzieri, G., Castelli, F., & Vivoni, E. R. (2011). Vegetation Dynamics within the North American Monsoon Region. *Journal of Climate*, *24*(6), 1763–1783. https://doi.org/10.1175/2010JCLI3847.1

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data*, *2*, 150066. https://doi.org/10.1038/sdata.2015.66

Gagne, D. J., McGovern, A., & Brotzge, J. (2009). Classification of Convective Areas Using Decision Trees. *Journal of Atmospheric and Oceanic Technology*, *26*(7), 1341–1353. https://doi.org/10.1175/2008JTECHA1205.1

Gagne, D. J., McGovern, A., Haupt, S. E., Sobash, R. A., Williams, J. K., & Xue, M. (2017). Storm-Based Probabilistic Hail Forecasting with Machine Learning Applied to Convection-Allowing Ensembles. *Weather and Forecasting*, *32*(5), 1819–1840. https://doi.org/10.1175/WAF-D-17-0010.1

Galaz, J., Cienfuegos, R., Echeverría, A., Pereira, S., Bertín, C., Prato, G., & Karich, J. C. (2022). Integrating tsunami simulations in web applications using BROWNI, an open source client-side GPU-powered tsunami simulation library. *Computers & Geosciences*, *159*, 104976. https://doi.org/10.1016/j.cageo.2021.104976

Géron, A. (2019). *Hands-On machine learning with scikit-learn, Keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems* (Second edition). O'Reilly.

Goodman, S. J., Blakeslee, R. J., Koshak, W. J., Mach, D., Bailey, J., Buechler, D., Carey, L., Schultz, C., Bateman, M., McCaul, E., & Stano, G. (2013). The GOES-R Geostationary Lightning Mapper (GLM). *Atmospheric Research*, *125–126*, 34–49. https://doi.org/10.1016/j.atmosres.2013.01.006

Gravelle, C. M., Mecikalski, J. R., Line, W. E., Bedka, K. M., Petersen, R. A., Sieglaff, J. M., Stano, G. T., & Goodman, S. J. (2016). Demonstration of a GOES-R Satellite Convective Toolkit to "Bridge the Gap" between Severe Weather Watches and Warnings: An Example from the 20 May 2013 Moore, Oklahoma, Tornado Outbreak. *Bulletin of the American Meteorological Society*, *97*(1), 69–84. https://doi.org/10.1175/BAMS-D-14-00054.1

Green, W. H., & Ampt, G. A. (1911). Studies on Soil Phyics. *The Journal of Agricultural Science*, *4*(1), 1–24. https://doi.org/10.1017/S0021859600001441

Haan, C. T. (Charles T. (1994). *Design hydrology and sedimentology for small catchments*. Academic Press. http://archive.org/details/designhydrologys0000haan

Han, D., Lee, J., Im, J., Sim, S., Lee, S., & Han, H. (2019). A Novel Framework of Detecting Convective Initiation Combining Automated Sampling, Machine Learning, and Repeated Model Tuning from Geostationary Satellite Data. *Remote Sensing*, *11*(12), Article 12. https://doi.org/10.3390/rs11121454

Han, H., Lee, S., Im, J., Kim, M., Lee, M.-I., Ahn, M. H., & Chung, S.-R. (2015). Detection of Convective Initiation Using Meteorological Imager Onboard Communication, Ocean, and Meteorological Satellite Based on Machine Learning Approaches. *Remote Sensing*, *7*(7), Article 7. https://doi.org/10.3390/rs70709184

Han, L., Sun, J., Zhang, W., Xiu, Y., Feng, H., & Lin, Y. (2017). A machine learning nowcasting method based on real-time reanalysis data. *Journal of Geophysical Research: Atmospheres*, *122*(7), 4038–4051. https://doi.org/10.1002/2016JD025783

Harjupa, W., Abdillah, M. R., Azura, A., Putranto, M. F., Marzuki, M., Nauval, F., Risyanto, Saufina, E., Jumianti, N., & Fathrio, I. (2022). On the utilization of RDCA method for detecting and predicting the occurrence of heavy rainfall in

Indonesia. *Remote Sensing Applications: Society and Environment*, *25*, 100681. https://doi.org/10.1016/j.rsase.2021.100681
Hasan, B. M. S., & Abdulazeez, A. M. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, *2*(1), Article 1.
Hayward, L., Whitworth, M., Pepin, N., & Dorling, S. (2020). Review article: A comprehensive review of datasets and methodologies employed to produce thunderstorm climatologies. *Natural Hazards and Earth System Sciences*, *20*(9), 2463–2482. https://doi.org/10.5194/nhess-20-2463-2020
Henderson, D. S., Otkin, J. A., & Mecikalski, J. R. (2021). Evaluating Convective Initiation in High-Resolution Numerical Weather Prediction Models Using GOES-16 Infrared Brightness Temperatures. *Monthly Weather Review*, *149*(4), 1153–1172. https://doi.org/10.1175/MWR-D-20-0272.1
Hilburn, K. A., Ebert-Uphoff, I., & Miller, S. D. (2020). Development and Interpretation of a Neural-Network-Based Synthetic Radar Reflectivity Estimator Using GOES-R Satellite Observations. *Journal of Applied Meteorology and Climatology*, *60*(1), 3–21. https://doi.org/10.1175/JAMC-D-20-0084.1
Holle, R. L., & Murphy, M. J. (2015). Lightning in the North American Monsoon: An Exploratory Climatology. *Monthly Weather Review*, *143*(5), 1970–1977. https://doi.org/10.1175/MWR-D-14-00363.1
Horton, R. (1933). The Rôle of infiltration in the hydrologic cycle. *Eos, Transactions American Geophysical Union*, *14*(1), 446–460. https://doi.org/10.1029/TR014i001p00446
Hosmer, D., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (Third). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118548387
Huff, F. A. (1967). Time distribution of rainfall in heavy storms. *Water Resources Research*, *3*(4), 1007–1019. https://doi.org/10.1029/WR003i004p01007
IMPLAN. (2012). *Atlas de Riesgos Naturales de la ciudad de Los Mochis 2011* (p. 109). Instituto Municipal de Planeación de Ahome. http://rmgir.proyectomesoamerica.org/PDFMunicipales/2011/vr_25001_AR_AHOME.pdf
INEGI. (2008). *Conjunto de datos vectoriales escala 1:1 000 000. Unidades climáticas* [[Formato Shapefile]]. Instituto Nacional de Estadística, Geografía e Informática.
INEGI. (2013). *Continuo de Elevaciones Mexicano 3.0 (CEM 3.0)* [[Modelo Digital de Elevación]]. Instituto Nacional de Estadística, Geografía e Informática. http://www.beta.inegi.org.mx/app/geo2/elevacionesmex/
INEGI. (2017). *Modelo digital de elevación de alta resolución Lidar* (1°) [[Modelo Digital de Elevación]]. Instituto Nacional de Estadística Geografía e Informática. http://www.inegi.org.mx/
INIFAP. (2018). *Red Nacional de Estaciones Agrometeorológicas Automatizadas. Instituto Nacional de investigaciones Forestales y Agropecuarias*. [Base de datos]. https://clima.inifap.gob.mx/lnmysr/Estaciones/MapaEstaciones
INIFAP & CONABIO. (2001). *Edafología* (Instituto Nacional de investigaciones Forestales y Agropecuarias) [[Formato Shapefile]]. http://www.conabio.gob.mx/informacion/gis/?vns=eda251mgw

Jergensen, G. E., McGovern, A., Lagerquist, R., & Smith, T. (2020). Classifying Convective Storms Using Machine Learning. *Weather and Forecasting*, *35*(2), 537–559. https://doi.org/10.1175/WAF-D-19-0170.1
Jiang, P., Li, C., Li, R., & Yang, H. (2019). An innovative hybrid air pollution early-warning system based on pollutants forecasting and Extenics evaluation. *Knowledge-Based Systems*, *164*, 174–192. https://doi.org/10.1016/j.knosys.2018.10.036
Kete, S. C. R., Suprihatin, Tarigan, S. D., & Effendi, H. (2019). Land use classification based on object and pixel using Landsat 8 OLI in Kendari City, Southeast Sulawesi Province, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *284*(1), 012019. https://doi.org/10.1088/1755-1315/284/1/012019
Khlopenkov, K. V., Bedka, K. M., Cooney, J. W., & Itterly, K. (2021). Recent Advances in Detection of Overshooting Cloud Tops From Longwave Infrared Satellite Imagery. *Journal of Geophysical Research: Atmospheres*, *126*(14), e2020JD034359. https://doi.org/10.1029/2020JD034359
Kim, M., Im, J., Park, H., Park, S., Lee, M.-I., & Ahn, M.-H. (2017). Detection of Tropical Overshooting Cloud Tops Using Himawari-8 Imagery. *Remote Sensing*, *9*(7), Article 7. https://doi.org/10.3390/rs9070685
Kim, M., Lee, J., & Im, J. (2018). Deep learning-based monitoring of overshooting cloud tops from geostationary satellite data. *GIScience & Remote Sensing*, *55*(5), 763–792. https://doi.org/10.1080/15481603.2018.1457201
Krinitskiy, M., Sprygin, A., Elizarov, S., Narizhnaya, A., Shikhov, A., & Chernokulsky, A. (2023). Towards the Accurate Automatic Detection of Mesoscale Convective Systems in Remote Sensing Data: From Data Mining to Deep Learning Models and Their Applications. *Remote Sensing*, *15*(14), Article 14. https://doi.org/10.3390/rs15143493
La Fata, A., Amato, F., Bernardi, M., D'Andrea, M., Procopio, R., & Fiori, E. (2021). Cloud-to-Ground lightning nowcasting using Machine Learning. *2021 35th International Conference on Lightning Protection (ICLP) and XVI International Symposium on Lightning Protection (SIPDA)*, *1*, 1–6. https://doi.org/10.1109/ICLPandSIPDA54065.2021.9627428
La Fata, A., Amato, F., Bernardi, M., D'Andrea, M., Procopio, R., & Fiori, E. (2022). Horizontal grid spacing comparison among Random Forest algorithms to nowcast Cloud-to-Ground lightning occurrence. *Stochastic Environmental Research and Risk Assessment*, *36*(8), 2195–2206. https://doi.org/10.1007/s00477-022-02222-1
Lagerquist, R., Stewart, J. Q., Ebert-Uphoff, I., & Kumler, C. (2021). Using Deep Learning to Nowcast the Spatial Coverage of Convection from Himawari-8 Satellite Data. *Monthly Weather Review*, *149*(12), 3897–3921. https://doi.org/10.1175/MWR-D-21-0096.1
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791
Lee, S., Han, H., Im, J., Jang, E., & Lee, M.-I. (2017). Detection of deterministic and probabilistic convection initiation using Himawari-8 Advanced Himawari

Imager data. *Atmospheric Measurement Techniques*, *10*(5), 1859–1874. https://doi.org/10.5194/amt-10-1859-2017

Lee, Y., Han, D., Ahn, M.-H., Im, J., & Lee, S. J. (2019). Retrieval of Total Precipitable Water from Himawari-8 AHI Data: A Comparison of Random Forest, Extreme Gradient Boosting, and Deep Neural Network. *Remote Sensing*, *11*(15), Article 15. https://doi.org/10.3390/rs11151741

Lee, Y., Kummerow, C. D., & Ebert-Uphoff, I. (2021). Applying machine learning methods to detect convection using Geostationary Operational Environmental Satellite-16 (GOES-16) advanced baseline imager (ABI) data. *Atmospheric Measurement Techniques*, *14*(4), 2699–2716. https://doi.org/10.5194/amt-14-2699-2021

Leinonen, J., Hamann, U., & Germann, U. (2022). Seamless Lightning Nowcasting with Recurrent-Convolutional Deep Learning. *Artificial Intelligence for the Earth Systems*, *1*(4). https://doi.org/10.1175/AIES-D-22-0043.1

Lenk, S., Senf, F., Deneke, H., & Hungershoefer, K. (2018). *On the sensitivity of nowcasting algorithms for convective initiation in satellite images to automatic tracking techniques*. https://www.semanticscholar.org/paper/ON-THE-SENSITIVITY-OF-NOWCASTING-ALGORITHMS-FOR-IN-Lenk-Senf/0328bbc3a16636e709ce6fd282fa99140c742c72

León-Cruz, J. F., Caetano, E., Cortés-Ramos, J., Dominguez, C., & Méndez-Pérez, J. M. (2023). Thunderstorm and hailstorm environments in Mexico. *International Journal of Climatology*, *43*(10), 4379–4395. https://doi.org/10.1002/joc.8093

León-Cruz, J. F., Carbajal Henken, C., Carbajal, N., & Fischer, J. (2021). Spatio-Temporal Distribution of Deep Convection Observed along the Trans-Mexican Volcanic Belt. *Remote Sensing*, *13*(6), Article 6. https://doi.org/10.3390/rs13061215

Lindstrom, S. (2017, December 18). Changes to the GOES-16 CONUS sector. *CIMSS Satellite Blog*. https://cimss.ssec.wisc.edu/satellite-blog/archives/26386

Liu, Q., Li, Y., Yu, M., Chiu, L. S., Hao, X., Duffy, D. Q., & Yang, C. (2019). Daytime Rainy Cloud Detection and Convective Precipitation Delineation Based on a Deep Neural Network Method Using GOES-16 ABI Images. *Remote Sensing*, *11*(21), Article 21. https://doi.org/10.3390/rs11212555

Machado, L. a. T., & Rossow, W. B. (1993). Structural Characteristics and Radiative Properties of Tropical Cloud Clusters. *Monthly Weather Review*, *121*(12), 3234–3260. https://doi.org/10.1175/1520-0493(1993)121<3234:SCARPO>2.0.CO;2

Maddox, R. A. (1980). Mesoscale Convective Complexes. *Bulletin of the American Meteorological Society*, *61*(11), 1374–1387.

Mahomed, M., Clulow, A. D., Strydom, S., Mabhaudhi, T., & Savage, M. J. (2021). Assessment of a Ground-Based Lightning Detection and Near-Real-Time Warning System in the Rural Community of Swayimane, KwaZulu-Natal, South Africa. *Weather, Climate, and Society*, *13*(3), 605–621. https://doi.org/10.1175/WCAS-D-20-0116.1

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

McGovern, A., Chase, R. J., Flora, M., Gagne, D. J., Lagerquist, R., Potvin, C. K., Snook, N., & Loken, E. (2023). A Review of Machine Learning for Convective Weather. *Artificial Intelligence for the Earth Systems*, *1*(aop), 1–61. https://doi.org/10.1175/AIES-D-22-0077.1

Mecikalski, J. R., & Bedka, K. M. (2006). Forecasting Convective Initiation by Monitoring the Evolution of Moving Cumulus in Daytime GOES Imagery. *Monthly Weather Review*, *134*(1), 49–78. https://doi.org/10.1175/MWR3062.1

Mecikalski, J. R., Williams, J. K., Jewett, C. P., Ahijevych, D., LeRoy, A., & Walker, J. R. (2015). Probabilistic 0–1-h Convective Initiation Nowcasts that Combine Geostationary Satellite Observations and Numerical Weather Prediction Model Data. *Journal of Applied Meteorology and Climatology*, *54*(5), 1039–1059. https://doi.org/10.1175/JAMC-D-14-0129.1

Mejia, J. F., & Douglas, M. W. (2009). *Relationships between Moisture Surges and Mesoscale- to Large-Scale Convection from Multi-year Satellite Imagery and North American Regional Reanalysis Data*. *2009*, A51F-0170.

Ministerio de FOMENTO. (2016). *Tabla 2.3 Norma 5.2-IC drenaje superficial*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://hidrologia.usal.es/temas/Hid_Sup_3.pdf

Mueller, C. K., Wilson, J. W., & Crook, N. A. (1993). The Utility of Sounding and Mesonet Data to Nowcast Thunderstorm Initiation. *Weather and Forecasting*, *8*(1), 132–146. https://doi.org/10.1175/1520-0434(1993)008<0132:TUOSAM>2.0.CO;2

Negrete Correa, D. U. (2016). *Desarrollo de un modelo numérico para la simulación de procesos hidráulicos* [masterThesis, Instituto Potosino de Investigación Científica y Tecnológica, A.C.]. https://repositorio.ipicyt.edu.mx///handle/11627/2930

Nguyen, P., Shearer, E. J., Tran, H., Ombadi, M., Hayatbini, N., Palacios, T., Huynh, P., Braithwaite, D., Updegraff, G., Hsu, K., Kuligowski, B., Logan, W. S., & Sorooshian, S. (2019). The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. *Scientific Data*, *6*(1), Article 1. https://doi.org/10.1038/sdata.2018.296

NOAA & NASA program. (2019). *GOES - R. Website* [Official Website]. Geostationary Operational Environmental Satellites - R Series. https://www.goes-r.gov/

Pereira-Nunes, M., França, G. B., & de Almeida, V. A. (2024). Application of Machine Learning Models for Convective Meteorological Events. *Pure and Applied Geophysics*. https://doi.org/10.1007/s00024-023-03395-8

Petersen, R., Aune, R., & Rink, T. (2010). *Objective short-range forecasts of the pre-convective environment using SEVIRI data*. 2010 EUMETSAT Meteorological Satellite Conf., Cordoba, Spain. https://www.eumetsat.int/media/7208

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., Arnold, G. T., Zhang, Z., Hubanks, P. A., Holz, R. E., Yang, P., Ridgway, W.

L., & Riedi, J. (2017). The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), 502–525. https://doi.org/10.1109/TGRS.2016.2610522

Prein, A. F., Rasmussen, R. M., Wang, D., & Giangrande, S. E. (2021). Sensitivity of organized convective storms to model grid spacing in current and future climates. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2195), 20190546. https://doi.org/10.1098/rsta.2019.0546

Qing, Z., Zeng, Q., Wang, H., Liu, Y., Xiong, T., & Zhang, S. (2022). ADASYN-LOF Algorithm for Imbalanced Tornado Samples. *Atmosphere*, *13*(4), Article 4. https://doi.org/10.3390/atmos13040544

Ramirez, S., & Lizarazo, I. (2017). Detecting and tracking mesoscale precipitating objects using machine learning algorithms. *International Journal of Remote Sensing*, *38*(18), 5045–5068. https://doi.org/10.1080/01431161.2017.1323280

Ramírez-López, A. K., Carbajal, N., Pineda-Martínez, L. F., & Tuxpan-Vargas, J. (2023). Cloud climatology of northwestern Mexico based on MODIS data. *European Journal of Remote Sensing*, *56*(1), 2278066. https://doi.org/10.1080/22797254.2023.2278066

Ramos-Pérez, O., Adams, D. K., Ochoa-Moya, C. A., & Quintanar, A. I. (2022). A Climatology of Mesoscale Convective Systems in Northwest Mexico during the North American Monsoon. *Atmosphere*, *13*(5), Article 5. https://doi.org/10.3390/atmos13050665

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), Article 7743. https://doi.org/10.1038/s41586-019-0912-1

Ritter, J., Berenguer, M., Corral, C., Park, S., & Sempere-Torres, D. (2020). ReAFFIRM: Real-time Assessment of Flash Flood Impacts – a Regional high-resolution Method. *Environment International*, *136*, 105375. https://doi.org/10.1016/j.envint.2019.105375

Roberts, R. D., & Rutledge, S. (2003). Nowcasting Storm Initiation and Growth Using GOES-8 and WSR-88D Data. *Weather and Forecasting*, *18*(4), 562–584. https://doi.org/10.1175/1520-0434(2003)018<0562:NSIAGU>2.0.CO;2

Romero Lankao, P. (2010). Water in Mexico City: What will climate change bring to its history of water-related hazards and vulnerabilities? *Environment and Urbanization*, *22*(1), 157–178. https://doi.org/10.1177/0956247809362636

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Rossow, W. B., & Schiffer, R. A. (1991). ISCCP Cloud Data Products. *Bulletin of the American Meteorological Society*, *72*(1), 2–20. https://doi.org/10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2

Rudlosky, S. D., Goodman, S. J., Virts, K. S., & Bruning, E. C. (2019). Initial Geostationary Lightning Mapper Observations. *Geophysical Research Letters*, *46*(2), 1097–1104. https://doi.org/10.1029/2018GL081052

Rutledge, S. A., Hilburn, K. A., Clayton, A., Fuchs, B., & Miller, S. D. (2020). Evaluating Geostationary Lightning Mapper Flash Rates Within Intense Convective Storms. *Journal of Geophysical Research: Atmospheres*, *125*(14), e2020JD032827. https://doi.org/10.1029/2020JD032827

Sahour, H., Gholami, V., Torkaman, J., Vazifedan, M., & Saeedi, S. (2021). Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings. *Environmental Earth Sciences*, *80*(22), 747. https://doi.org/10.1007/s12665-021-10054-5

Salgado Pilario, K. E., Cao, Y., & Shafiee, M. (2021). A Kernel Design Approach to Improve Kernel Subspace Identification. *IEEE Transactions on Industrial Electronics*, *68*(7), 6171–6180. https://doi.org/10.1109/TIE.2020.2996142

Samsi, S., Mattioli, C. J., & Veillette, M. S. (2019). Distributed Deep Learning for Precipitation Nowcasting. *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, 1–7. https://doi.org/10.1109/HPEC.2019.8916416

Sangiorgio, M., Barindelli, S., Guglieri, V., Biondi, R., Solazzo, E., Realini, E., Venuti, G., & Guariso, G. (2020). A Comparative Study on Machine Learning Techniques for Intense Convective Rainfall Events Forecasting. In O. Valenzuela, F. Rojas, L. J. Herrera, H. Pomares, & I. Rojas (Eds.), *Theory and Applications of Time Series Analysis* (pp. 305–317). Springer International Publishing. https://doi.org/10.1007/978-3-030-56219-9_20

Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., & Lebair, W. J. (2017). A Closer Look at the ABI on the GOES-R Series. *Bulletin of the American Meteorological Society*, *98*(4), 681–698. https://doi.org/10.1175/BAMS-D-15-00230.1

SEMARNAT. (2017, October 2). *Salvar vidas, objetivo de los sistemas de alerta temprana.* [Secretaría de Medio Ambiente y Recursos Naturales]. https://www.gob.mx/semarnat/articulos/salvar-vidas-objetivo-de-los-sistemas-de-alerta-temprana

Shenk, W. E. (1974). Cloud Top Height Variability of Strong Convective Cells. *Journal of Applied Meteorology and Climatology*, *13*(8), 917–922. https://doi.org/10.1175/1520-0450(1974)013<0917:CTHVOS>2.0.CO;2

SIATL-INEGI. (2016). *Simulador de Flujos de Agua de Cuencas Hidrográficas.* [Software Online]. Instituto Nacional de Estadística Geografía e Informática. http://antares.inegi.org.mx/analisis/red_hidro/SIATL/#

Siqueira, J. R., Marques, V. da S., Siqueira, J. R., & Marques, V. da S. (2021). Tracking and short-term forecasting of mesoscale convective cloud clusters over southeast Brazil using satellite infrared imagery. *Journal of Southern Hemisphere Earth Systems Science*, *71*(1), 1–16. https://doi.org/10.1071/ES19050

SMN, & CONAGUA. (2023). *Estaciones Meteorológicas Automáticas (EMAS). Servicio Meteorológico Nacional*. [Base de datos]. https://smn.conagua.gob.mx/es/observando-el-tiempo/estaciones-meteorologicas-automaticas-ema-s

Snodgrass, E. R., Girolamo, L. D., & Rauber, R. M. (2009). Precipitation Characteristics of Trade Wind Clouds during RICO Derived from Radar, Satellite, and Aircraft Measurements. *Journal of Applied Meteorology and Climatology*, *48*(3), 464–483. https://doi.org/10.1175/2008JAMC1946.1

Sobash, R. A., Romine, G. S., & Schwartz, C. S. (2020). A Comparison of Neural-Network and Surrogate-Severe Probabilistic Convective Hazard Guidance Derived from a Convection-Allowing Model. *Weather and Forecasting*, *35*(5), 1981–2000. https://doi.org/10.1175/WAF-D-20-0036.1

Steinkruger, D., Markowski, P., & Young, G. (2020). An Artificially Intelligent System for the Automated Issuance of Tornado Warnings in Simulated Convective Storms. *Weather and Forecasting*, *35*(5), 1939–1965. https://doi.org/10.1175/WAF-D-19-0249.1

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software*, *90*, 201–216. https://doi.org/10.1016/j.envsoft.2017.01.006

Tippett, M. K., Allen, J. T., Gensini, V. A., & Brooks, H. E. (2015). Climate and Hazardous Convective Weather. *Current Climate Change Reports*, *1*(2), 60–73. https://doi.org/10.1007/s40641-015-0006-6

Torres, R., Ohashi, O., & Pessin, G. (2019). A Machine-Learning Approach to Distinguish Passengers and Drivers Reading While Driving. *Sensors (Basel, Switzerland)*, *19*(14), 3174. https://doi.org/10.3390/s19143174

Tracy, C., & Mecikalski, J. R. (2023). Mesoscale Influences of Land Use, Topography, Antecedent Rainfall, and Atmospheric Conditions on Summertime Convective Storm Initiation under Weak Synoptic-Scale Forcing. *Weather and Forecasting*, *39*(1), 55–74. https://doi.org/10.1175/WAF-D-22-0216.1

Trafalis, T. B., Ince, H., & Richman, M. B. (2003). Tornado Detection with Support Vector Machines. In P. M. A. Sloot, D. Abramson, A. V. Bogdanov, Y. E. Gorbachev, J. J. Dongarra, & A. Y. Zomaya (Eds.), *Computational Science—ICCS 2003* (pp. 289–298). Springer. https://doi.org/10.1007/3-540-44864-0_30

Ukkonen, P., & Mäkelä, A. (2019). Evaluation of Machine Learning Classifiers for Predicting Deep Convection. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1784–1802. https://doi.org/10.1029/2018MS001561

United Nations. (2022). *Early Warnings for All* [United Nations. Climate Action]. https://www.un.org/en/climatechange/early-warnings-for-all#:~:text=The%20Action%20Plan%20calls%20for,and%20preparedness%20and%20response%20capabilities.

Vapnik, V. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, *24*, 774–780.

Vargas, N., & Magaña, V. (2020). Climatic risk in the Mexico city metropolitan area due to urbanization. *Urban Climate*, *33*, 100644. https://doi.org/10.1016/j.uclim.2020.100644

Veillette, M. S. (2013, January 9). *Convective Initiation Forecasts Through the Use of Machine Learning Methods*. 93rd American Meteorological Society Annual

Meeting. https://ams.confex.com/ams/93Annual/webprogram/Paper221560.html
Walker, J. R., MacKenzie, W. M., Mecikalski, J. R., & Jewett, C. P. (2012). An Enhanced Geostationary Satellite–Based Convective Initiation Algorithm for 0–2-h Nowcasting with Object Tracking. *Journal of Applied Meteorology and Climatology*, *51*(11), 1931–1949. https://doi.org/10.1175/JAMC-D-11-0246.1
Wanielista, M. P., & Yousef, Y. A. (1992). *Stormwater management*. Wiley-Interscience.
Wannachai, A., Aramkul, S., Suntaranont, B., Somchit, Y., & Champrasert, P. (2022). HERO: Hybrid Effortless Resilient Operation Stations for Flash Flood Early Warning Systems. *Sensors*, *22*(11), Article 11. https://doi.org/10.3390/s22114108
Weckwerth, T. M., & Parsons, D. B. (2006). A Review of Convection Initiation and Motivation for IHOP_2002. *Monthly Weather Review*, *134*(1), 5–22. https://doi.org/10.1175/MWR3067.1
WMO. (2022). *WMO Bulletin: Early Warning and Anticipatory Action*.
Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1
World Meteorological Organization. (2021). *WMO Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2019)*. https://library.wmo.int/records/item/57564-wmo-atlas-of-mortality-and-economic-losses-from-weather-climate-and-water-extremes-1970-2019#.YUg7QSviuUk
Yao, H., Li, X., Pang, H., Sheng, L., & Wang, W. (2020). Application of random forest algorithm in hail forecasting over Shandong Peninsula. *Atmospheric Research*, *244*, 105093. https://doi.org/10.1016/j.atmosres.2020.105093
Yuan, W., Tu, X., Su, C., Liu, M., Yan, D., & Wu, Z. (2021). Research on the Critical Rainfall of Flash Floods in Small Watersheds Based on the Design of Characteristic Rainfall Patterns. *Water Resources Management*, *35*(10), 3297–3319. https://doi.org/10.1007/s11269-021-02893-5
Yuan, X., Wang, L., Wu, P., Ji, P., Sheffield, J., & Zhang, M. (2019). Anthropogenic shift towards higher risk of flash drought over China. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-12692-7
Yucel, I., & Onen, A. (2014). Evaluating a mesoscale atmosphere model and a satellite-based algorithm in estimating extreme rainfall events in northwestern Turkey. *Natural Hazards and Earth System Sciences*, *14*(3), 611–624. https://doi.org/10.5194/nhess-14-611-2014
Zhang, W., Han, L., Sun, J., Guo, H., & Dai, J. (2019). Application of Multi-channel 3D-cube Successive Convolution Network for Convective Storm Nowcasting. *2019 IEEE International Conference on Big Data (Big Data)*, 1705–1710. https://doi.org/10.1109/BigData47090.2019.9005568
Zheng, Z., Wang, J., Shi, L., Zhao, S., Hou, J., Sun, L., & Dong, L. (2022). Generating phone-quality records to train machine learning models for smartphone-based earthquake early warning. *Journal of Seismology*, *26*(3), 439–454. https://doi.org/10.1007/s10950-021-10058-5
Zhou, K., Zheng, Y., Dong, W., & Wang, T. (2020). A Deep Learning Network for Cloud-to-Ground Lightning Nowcasting with Multisource Data. *Journal of*

Atmospheric and Oceanic Technology, *37*(5), 927–942. https://doi.org/10.1175/JTECH-D-19-0146.1

Annexes

Annex 1.
Dávila-Ortiz, R., Carbajal-Pérez, J. N., Velázquez-Zapata, J. A., & Tuxpan-Vargas, J. (2024). Approximation of a Convective-Event-Monitoring System Using GOES-R Data and Ensemble ML Models. *Remote Sensing*, 16(4), 675.

Annex 2.
Graphical Abstract of the paper: Approximation of a Convective-Event-Monitoring System Using GOES-R Data and Ensemble ML Models.

Annex 3.
Dávila Ortiz, R., Tuxpan Vargas, J., & Velázquez Zapata, J. A. (2023). Identification of Deep Convection Clouds Using ABIGOES Data and Machine Learning Techniques: The Case of Los Mochis, Sinaloa, Mexico. 2023 Mexican International Conference on Computer Science (ENC), 1–7. https://doi.org/10.1109/ENC60556.2023.10508667

Annex 4.
Dávila Ortiz, R., Tuxpan Vargas, J., & Velázquez Zapata, J. A. (in press). La producción del riesgo ante eventos de inundación en Los Mochis, Sinaloa. In Riesgos y desastres relacionados con agua transformación del territorio, inundaciones y contaminación. El Colegio de San Luis, A.C.

Annex 5.
Dávila Ortiz, R., & Tuxpan Vargas, J. (2024, enero 11). Cazando nubes de tormenta. Propuesta para un sistema de monitoreo y alerta temprana basado en inteligencia artificial en México. Opinión Pública SLP. https://opslp.mx/cazando-nubes-de-tormenta-propuesta-para-un-sistema-de-monitoreo-y-alerta-temprana-basado-en-inteligencia-artificial-en-mexico/

remote sensing

MDPI

Article

Approximation of a Convective-Event-Monitoring System Using GOES-R Data and Ensemble ML Models

Rodrigo Dávila-Ortiz [1], José Noel Carbajal-Pérez [1], Juan Alberto Velázquez-Zapata [2] and José Tuxpan-Vargas [3,*]

[1] Instituto Potosino de Investigación Científica y Tecnológica, A.C. División de Geociencias Aplicadas, Camino a la Presa San José 2055, San Luis Potosí, Mexico; rodrigo.davila@ipicyt.edu.mx (R.D.-O.); noelc@ipicyt.edu.mx (J.N.C.-P.)

[2] CONAHCYT-El Colegio de San Luis, A.C., Parque Macul 155, San Luis Potosí, Mexico; alberto.velazquez@colsan.edu.mx

[3] CONAHCYT-Instituto Potosino de Investigación Científica y Tecnológica, A.C. División de Geociencias Aplicadas, Camino a la Presa San José 2055, San Luis Potosí, Mexico

* Correspondence: jose.tuxpan@ipicyt.edu.mx; Tel.: +52-(444)-8342-000 (ext. 3208)

Abstract: The presence of deep convective clouds is directly related to potential convective hazards, such as lightning strikes, hail, severe storms, flash floods, and tornadoes. On the other hand, Mexico has a limited and heterogeneous network of instruments that allow for efficient and reliable monitoring and forecasting of such events. In this study, a quasi-real-time framework for deep convective cloud identification and modeling based on machine learning (ML) models was developed. Eight different ML models and model assembly approaches were fed with Interest Fields estimated from Advanced Baseline Imager (ABI) sensor data on the Geostationary Operational Environmental Satellite-R Series (GOES-R) for one region in central Mexico and another in northeastern Mexico, both selected for their intense convective activity and high levels of vulnerability to severe weather. The results indicate that a simple approach such as Logistic Regression (LR) or Random Forest (RF) can be a good alternative for the identification and simulation of deep convective clouds in both study areas, with a probability of detection of (POD) $\approx$ 0.84 for Los Mochis and POD of $\approx$ 0.72 for Mexico City. Similarly, the false alarm ratio (FAR) $\approx$ 0.2 and FAR $\approx$ 0.4 values were obtained for Los Mochis and Mexico City, respectively. Finally, a post-processing filter based on lightning incidence (Lightning Filter) was applied with data from the Geostationary Lightning Mapper (GLM) of the GOES-16 satellite, showed great potential to improve the probability of detection (POD) of the ML models. This work sets a precedent for the implementation of an early-warning system for hazards associated with intense convective activity in Mexico.

Keywords: ABI-GOES data; convective-hazard forecasting; deep convective clouds; machine learning; quasi-real-time framework

check for updates

Citation: Dávila-Ortiz, R.; Carbajal-Pérez, J.N.; Velázquez-Zapata, J.A.; Tuxpan-Vargas, J. Approximation of a Convective-Event-Monitoring System Using GOES-R Data and Ensemble ML Models. *Remote Sens.* **2024**, *16*, 675. https://doi.org/10.3390/rs16040675

Academic Editors: Carmine Serio and Xiefei Zhi

Received: 19 December 2023
Revised: 26 January 2024
Accepted: 8 February 2024
Published: 14 February 2024

1. Introduction

Convective hazards, which are characterized by dynamic atmospheric processes involving the vertical movement of air masses, pose significant challenges to human safety and infrastructure resilience. These hazards encompass a spectrum of high-impact phenomena such as thunderstorms, wind, hail, tornadoes, lightning, and floods [1–4]. Due to the rapid evolution and microscale variability of deep convection events, the task of forecasting and implementing early mitigation measures becomes inherently complex. In addition, many countries, like Mexico, lack reliable observation networks and short-range weather forecasting systems [5].

In this context, several studies have focused on the development of algorithms for convective initiation (CI) nowcasts, i.e., short-term weather forecasting that specifically focuses on predicting the initiation and development of deep moist convective activity [6]. Convective initiation nowcasts utilize various observational data, such as radar imagery (e.g., the Thunderstorm Identification, Tracking, Analysis, and Nowcasting; TITAN [7]), satellite data [1,8–10], atmospheric instability indices (e.g., [11]), numerical weather predictions (NWP, e.g., the Corridor Integrated Weather System; CIWS [12]), and other meteorological parameters to identify favorable conditions for the initiation of thunderstorms. For example, the Satellite Convection Analysis and Tracking (SATCAST [13]) is a CI nowcasting expert system that uses eight predictors, called "Interest Fields" based on infrared Geostationary Operational Environmental Satellite (GOES) data, to forecast CI with 0–1 h lead times [14]. In this work, CI is defined as the first detection of Weather Surveillance Radar-1988 Doppler (WSR-88D) reflectivities $\geq$ 35 dBZ produced by convective clouds and satellite-derived atmospheric motion vectors (AMVs) for tracking individual cumulus clouds. Subsequently, a second upgrade of this system, SATCAST version 2 (STACASTv2), was proposed by [15] and includes object-tracking approaches used to compute temporal changes in brightness temperature. Later additions include the integration of NWP data [12].

Recently, with the rise of ML approaches in the field of geosciences [16] and the availability of higher-quality satellite information, several research works based on the use of ML techniques for CI identification have emerged. For example, ref. [17] proposed a CI algorithm that combines Interest Fields derived from GOES-R and NWP model data. This study relied on Logistic Regression (LR) and Random Forest (RF) models that provided better predictions than previous approaches, in addition to probabilistic rather than binary classification predictions. Through experimentation, they validated the performance of these CI algorithms in the United States, with a FAR of 0.1–0.18 lower than existing deterministic CI detection algorithms for GOES (FAR $\approx$ 0.48–0.6 [15]).

In the related literature, several researches have reported the development and implementation of CI nowcast algorithms based on ML approaches for different geostationary satellites, such as Himawari-8 [3,18], the Communication, Ocean, and Meteorological Satellite (COMS [2]), and Meteosat [19]. Other forecasting systems based on satellite data were developed to aid in diagnosing the characteristics of the preconvective environment, and deep moist convection are the convective occurrence (CO) algorithms. CO approaches include modeling convective initiation/decay and advection of existing storms [14], the 0–9 h NearCast model, [9,20], object tracking techniques such as optical flow [21], and thunderstorm occurrence (TO [22–24]).

Overshooting cloud top (OT) detection approaches consist of algorithms designed to identify and analyze cloud formations, with a particular focus on overshooting cloud tops (also called anvil domes and are defined as domelike clouds forming above a cumulonimbus cloud top or penetrating the tropopause [25]). OT clouds form when strong updrafts in thunderstorms push cloud tops above their equilibrium height (level of neutral buoyancy). These OT cloud tops are often associated with severe weather events. OT cloud detection approaches can be based on cloud top temperature thresholds (e.g., [26,27]), but also on ML and Deep Learning (DL) approaches. For example, in [28], three machine learning techniques, RF, LR, and extremely randomized trees (ERT), were used to develop OT classification models, achieving an average POD = 0.77 and FAR = 0.36 with the RF model. In another study, a new approach for OT detection was proposed using a Convolutional Neural Network (CNN), reporting a POD = 0.79 and a FAR = 0.09 [29].

This study proposes a quasi-real-time framework for deep convective clouds identification and modeling based on Machine Learning (ML) models and ABI-GOES data. The objectives of the study are first, to develop a deep convective cloud identification framework based on three principles: the use of open-source software and open-access products, automation, and scalability; second, to compare different ML approaches to determine which is optimal in terms of performance and computational cost; third, to quantify and examine relative variable importance and model contribution; and finally, to integrate a post-processing filter based on lightning incidence to improve the ability to detect deep convection events. The previously analyzed algorithms and approaches for the prediction of intense convective events or the identification of deep convective clouds were designed to integrate multi-source meteorological information, especially from radars.

The proposed framework in this study presents an alternative for the implementation of early-warning systems in areas with high levels of vulnerability to convective hazards in the context of limited meteorological monitoring infrastructure.

The structure of this paper is as follows: The test zones are described in Section 2; Section 3 presents the data and methods. The main findings and discussion are described in Section 4. Finally, the conclusions are presented in Section 5.

2. Test Sites

2.1. Los Mochis, Sin

Los Mochis is the main city of Ahome municipality, situated in the northeastern region of the Mexican state of Sinaloa. This city is positioned within the North American monsoon (NAM) core domain, leading to a notable occurrence of convective activity [30–32] and a substantial increase in summer rainfall [33]. Ref. [34] claims that the precipitation distribution in this zone is mainly related to the deep convective cloud pattern.

Figure 1 illustrates the geographic location of Los Mochis, which is positioned between the Gulf of Mexico to the west and the Sierra Madre Occidental (SMO) to the east. The simulation area, comprising a grid of 50 $\times$ 50 ABI-GOES cells, is situated within the latitudinal range of 25.25° to 26.5° and the longitudinal range of $-110°$ to $-108°$. The topography of this region is primarily flat, but there are also mountain systems at the edge of the SMO. This region is categorized as a Hot Desert climate (BWh [35]), with a mean annual precipitation of 335 mm, primarily concentrated during a pronounced wet season spanning from July to September. The precipitation patterns are significantly influenced by large-scale climatic phenomena such as El Niño–Southern Oscillation (ENSO) and hurricanes [36].

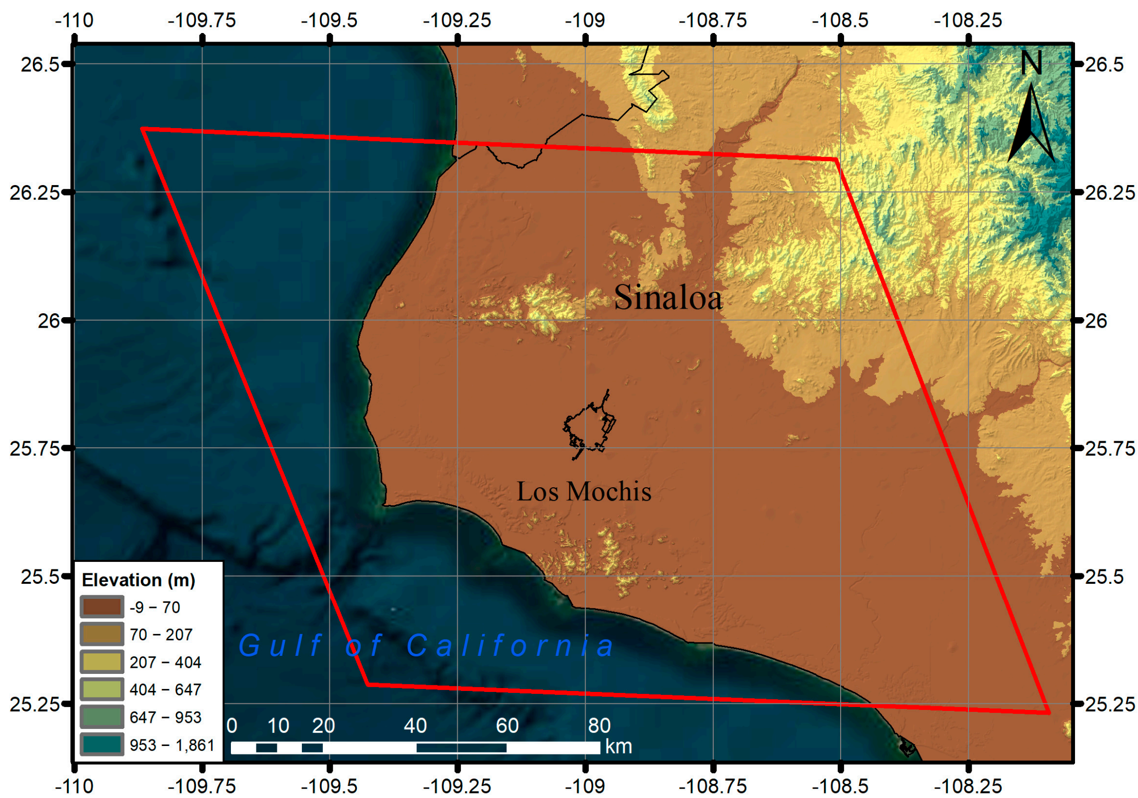


Figure 1. Land surface elevation over Los Mochis, [37]. The red border corresponds to the special simulation domain, which has a size of 50 $\times$ 50 cells.

2.2. Mexico City

Mexico City is Mexico's capital and largest city in the country and one of its main population centers. It is located in central Mexico, within the physiographic province called the Trans-Mexican Volcanic Belt, which crosses the country from the Pacific Ocean to the Gulf of Mexico, presenting a complex topography that ranges from 0 to 5000 m above sea level. The combination of different factors such as its latitudinal position, complex topography, the influence of tropical cyclones, cold fronts, and easterly waves [38], and even the effect of urbanization [39], cause this area to present intense convective activity. These conditions are a source of convective risk associated with the presence of tornadoes [38,40], floods [41], hailstorms, and thunderstorms [5]. In addition, socioeconomic conditions in this area cause a high degree of social vulnerability [42].

Figure 2 shows the geographic location of the Mexican City area, (the simulation spatial domain comprises the México and Morelos states) which is composed of a grid of 50 × 50 ABI-GOES cells and is situated within the north latitude range of 20° to 19° and the longitude range of −99.75° to −98.5°. The mean annual precipitation in Mexico City ranges between 621 and 1200 mm, and this zone presents a Subhumid Temperate climate type (Cw [35]).

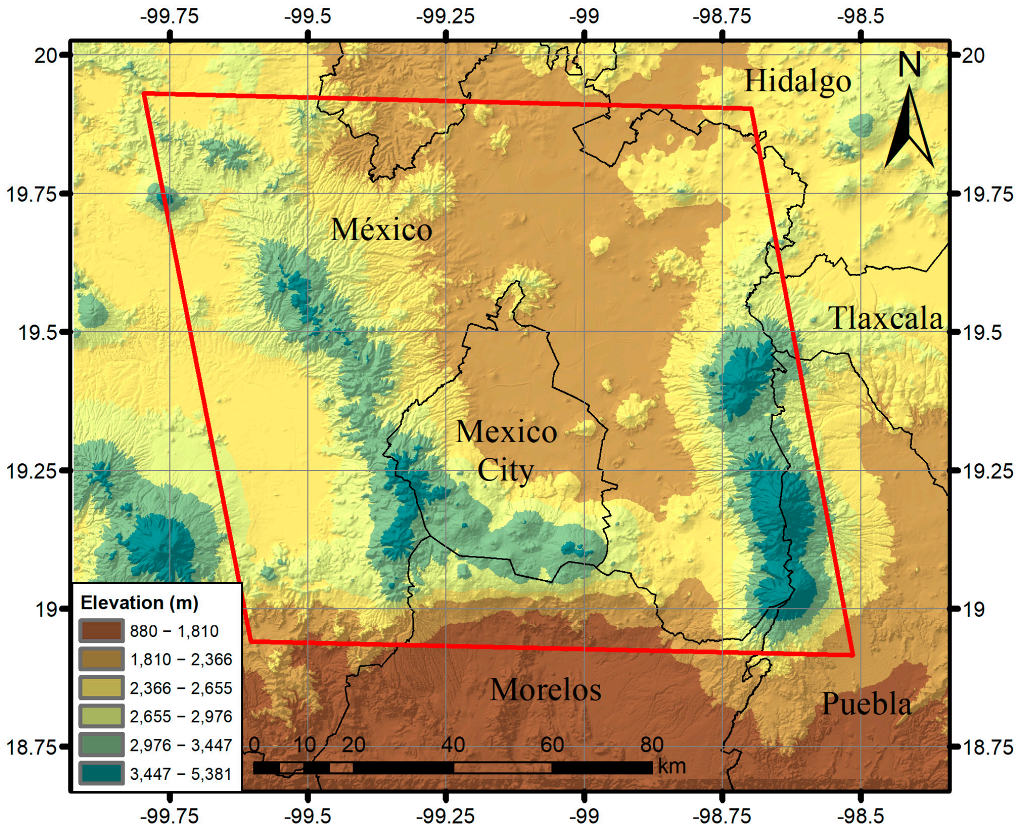


Figure 2. Land surface elevation over the study area called Mexico City (however, it also takes part of nearby territories [37]). The red border corresponds to the special simulation domain, which has a size of 50 × 50 cells.

3. Materials and Methods

3.1. Data

3.1.1. GOES-R Advanced Baseline Imager (ABI) Data

The GOES-R series of satellite instruments provide high-resolution and rapid-refresh observations of Earth's atmosphere and surface. This passive imaging radiometer has 16 spectral bands, including two visible channels, four near-infrared channels, and ten infrared channels [43]. Eight of the 16 ABI-GOES bands were used to generate 12 predictors based on the spectral properties of vertically developing clouds (Table 1).

Table 1. Specifications of ABI-GOES Bands.

ABI Bands	Bands Name	Center Wavelength (μm)	Temporal Resolution (min)	Best Spatial Resolution (km)
1	Blue	0.47	5	1
2	Red	0.64	5	0.5
3	Veggie	0.86	5	1
4	Cirrus	1.37	5	2
5	Snow/Ice	1.61	5	1
6	Cloud Particle Size	2.24	5	2
7	Shortwave Window	3.9	5	2
8	Upper-Level Tropospheric Water Vapor	6.19	5	2
9	Mid-Level Tropospheric Water Vapor	6.93	5	2
10	Lower-Level Water Vapor	7.37	5	2
11	Cloud-Top Phase	8.44	5	2
12	Ozone	9.61	5	2
13	Clean IR Longwave Window	10.33	5	2
14	IR Longwave Window	11.21	5	2
15	Dirty Longwave Window	12.29	5	2
16	CO_2 Longwave infrared	13.28	5	2

3.1.2. GOES-R Geostationary Lightning Mapper (GLM) Data

GLM is a single-channel, near-infrared optical transient detector that can detect momentary changes in an optical scene, indicating the presence of lightning [44]. The GLM sensor captures high-resolution images of lightning flashes, including cloud-to-ground, intra-cloud, and cloud-to-cloud lightning. It provides valuable data on lightning strikes' frequency, location, and intensity. A post-processing filter was generated using the presence of lightning as an indicator of convective activity, which usually occurs during the deep convection process.

3.1.3. Moderate-Resolution Imaging Spectroradiometer (MODIS) Data

MODIS is a passive imager mounted on both the Terra and Aqua sun-synchronous polar-orbiting satellites. This sensor provides pixel-level retrievals of cloud-top properties and cloud optical properties [45]. For target label generation, the variables "Cloud_Top_Preassure" and "Cloud_Optical_Thickness" (related to the optical and microphysical properties of deep convective clouds) from Collection 6 (C6) of MODIS products were collected. MODIS C6 introduces several improvements and refinements over the previous versions of the MODIS data products.

3.2. Methodology

3.2.1. Preparation of Training Dataset

Combinations of the brightness temperatures (Tbs) of the eight ABI bands were calculated to generate prediction variables called "Interest Fields" (Table 2). A wide variety of Interest Fields are used in deep convective cloud detection algorithms, adapted for various sensors, such as ABI-GOES [9,17], Advanced Himawari Imager (AHI [3,18]), Meteorological Imager (MI), a payload of the COMS [2], among others. These represent various physical properties of the cloud such as cloud-top temperature, cloud-top cooling rate, cloud optical depth, cloud-top height, etc.

Table 2. Summary of the input data used for the implementation of ML workflow and post-processing filter. The initial selection of predictor variables (Interest Fields) was taken from [3].

ID	Feature Names	Type	Sensor
CtT	Cloud-top temperature (11.2 μm TB)	Predictor	ABI
CtH01	Cloud-top height 01 (6.2–11.2 μm)	Predictor	ABI
CtH02	Cloud-top height 02 (6.2–7.3 μm)	Predictor	ABI
CtH03	Cloud-top height 03 (13.3–11.2 μm)	Predictor	ABI
CtG01	Cloud-top glaciation 01 (12.3–11.2 μm)	Predictor	ABI
CtG02	Cloud-top glaciation 02 (8.6–11.2 μm)	Predictor	ABI
CtG03	Cloud-top glaciation 03 (8.6–11.2 μm)–(11.2–12.3 μm)	Predictor	ABI
CtCrate	Cloud-top cooling rate (11.2 μm time trend)	Predictor	ABI
TChCtH01	Temporal changes in cloud-top height 01 (6.2–11.2 μm time trend)	Predictor	ABI
TChCtH02	Temporal changes in cloud-top height 02 (6.2–7.3 μm time trend)	Predictor	ABI
TChCtH03	Temporal changes in cloud-top height 03 (13.3–11.2 μm time trend)	Predictor	ABI
TChCtG03	Temporal changes in cloud-top glaciation ((8.6–11.2 μm)–(11.2–12.3 μm) time trend)	Predictor	ABI
LF	Lightning filter	Filter Array	GLM
CC_labels	Deep convective cloud labels	Target Variable	MODIS

The Interest Fields used as predictors in this study are similar to those in [3,18]. On the other hand, using the variables "Cloud_Optical_Thickness" (COT) and "cloud_top_pression_1km" (CtP) from MODIS C6, the set of Target Labels was built. As criteria for the labeling of Convective Cells (CC) within the MODIS survey, the well-established classification of the International Satellite Cloud Climatology Project (ISCCP; Ref. [46]) was selected, where cells with values of COT > 23 and CtP < 440 mb were considered as deep convective clouds.

For the construction of the reference dataset, all MODIS images available in the period 2018–2022 were analyzed between the months of greatest precipitation from May to September, for the two spatial domains of interest, Los Mochis and Mexico City. The MODIS images were cut according to the special domain of the study areas, and only sections with a percentage greater than 30% of labeled CC were selected to build the Target Label Dataset to avoid training the ML models with unbalanced class distribution data.

All sections were resampled with the pyresample function (python package for resampling geospatial image data) to homogenize them with the spatial domain of the Interest Fields, which has a size of 50 $\times$ 50 cells in both domains (Figures 1 and 2). The training dataset for Mexico City was generated with 24 reference events (60,000 examples), while the test dataset had 31 events (77,500 examples). On the other hand, for Los Mochis, 6 (15,000 examples) and 8 events (20,000 examples) were used for the training and testing datasets, respectively.

3.2.2. Machine Learning Approaches

In this study, eight different ML approaches are compared (Table 3), such as LR, RF, and Multi-layer perceptron (MLP), which are widely used in convective-hazard forecasting (e.g., [14,17,22,47,48]), along with ML approaches based on Ensemble Learning techniques (i.e., a group of predictors, called an ensemble, are trained together to improve predictive ability; Figure 3). All ML models, assembly strategies, metric estimation, and preprocessing methods were taken from scikit-learn, a free software ML library for the Python programming language (https://scikit-learn.org/stable/, accessed on 5 December 2023).

Table 3. Machine Learning Approaches. This work compares the performance of ML approaches: LR, RF, MLP, Bagging, LRstacking, RFstacking, Hvoting, and Svoting. DT and SVM have been used as assembled predictors for Stacking and Voting.

Models and Ensemble Methods	Approach	Abbreviation
Logistic regression	Single model	LR
Decision tree	Single model	DT
Support vector machine	Single model	SVM
Multi-layer perceptron	Single model	MLP
Random Forest	Ensemble model	RF
Bagging (Logistic regression)	Ensemble model	Bagging
Stacking (Logistic regression)	Ensemble model	LRstacking
Stacking (Random Forest)	Ensemble model	RFstacking
Hard Voting	Ensemble model	Hvoting
Soft Voting	Ensemble model	Svoting

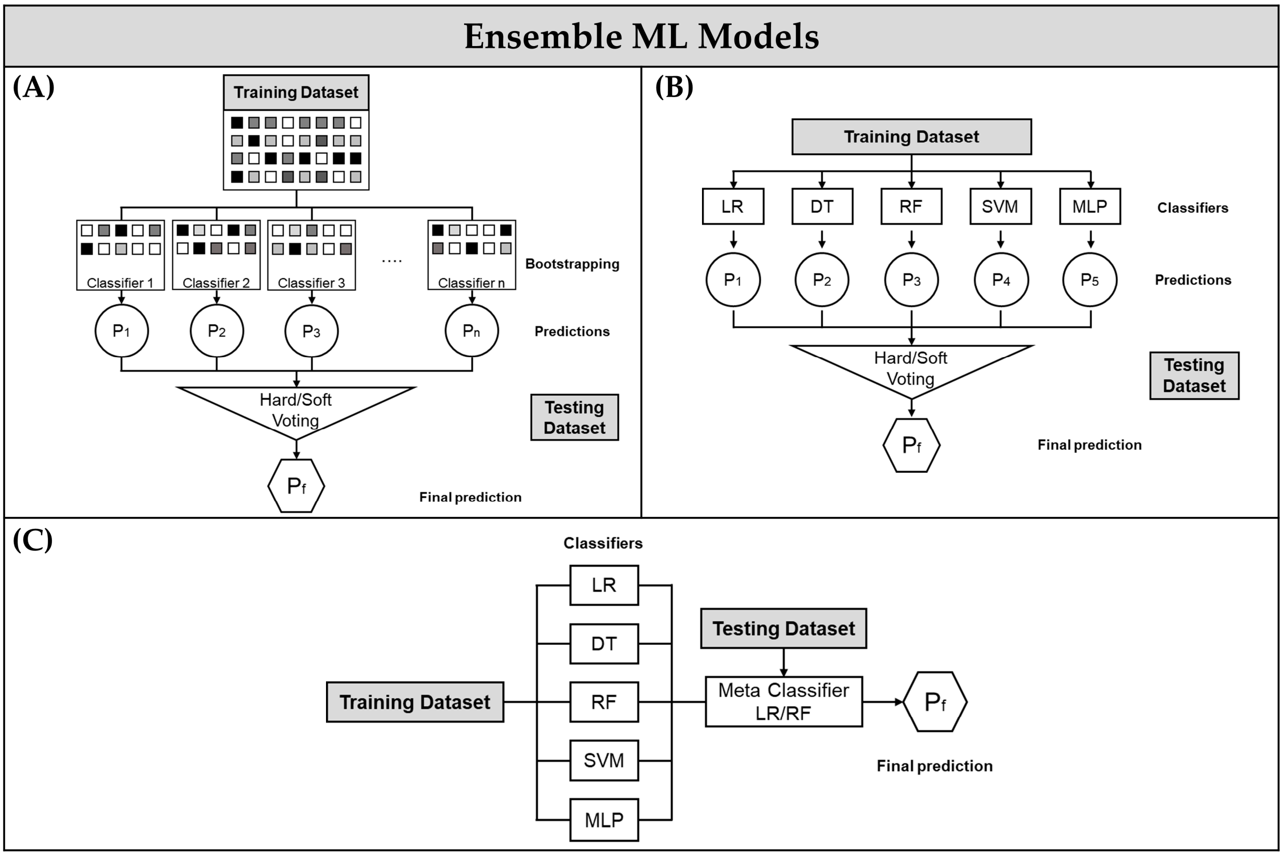


Figure 3. Diagram of ensemble ML methods. (**A**) Bagging, (**B**) Voting, (**C**) Stacking.

- Logistic regression

LR is a statistical model used to predict a binary variable (here, CC or noCC) from one or more independent variables or predictors [49]. The LR formula is given using

$$\mathrm{E}(Y) = \frac{1}{1 + \exp\left[-\left(\beta_0 + \sum_{j=\mathrm{i}}^{k} \beta_j X_j\right)\right]} \tag{1}$$

where E is the expected value of the dependent variable Y, which has values of 0 and 1, equivalent to noCC and CC, respectively; k is the number of predictors; and X_j is the value of the j th predictor. The parameters $\beta_0, \ldots, \beta_j$ are linear coefficients or weights for the predictor variables. Despite its simplicity, LR performs well in several real-world problems, including nonlinear problems in atmospheric science such as convection [23].

- Decision Tree

A DT is a non-parametric supervised learning method used for classification purposes. The goal is to generate a model that predicts the value of an independent variable by learning simple decision rules inferred from the data features [50]. Additional examples of convective-hazard forecasting with DT are the works of [2,51].

- Support Vector Machine

Developed by [52], SVM is a versatile supervised ML algorithm used for classification tasks. SVM works by finding an optimal hyperplane in a high-dimensional feature space to separate different classes of data points. The margin, which is the distance between the hyperplane and the nearest data points of each class (called support vectors), is found by minimizing the norm of the weight vector. Once the hyperplane is determined during the training phase, SVM can classify new, unseen data points by placing them on one side or the other of the decision boundary.

SVM is widely used in various convective-hazard forecasting applications, for example, in tornado prediction [53,54] or convective thunderstorm forecasting [47,55,56].

- Multi-layer perceptron

A MLP is a type of artificial feedforward neural network (ANN; the first conceptual model of an artificial neural network was developed by [57]) and consists of three kinds of layers of interconnected nodes called neurons. The layers that conform to the MLP have specific functions that allow the model to learn complex non-linear relationships between the input and output variables. These are the input layer, which receives the raw input data, so its number of neurons depends on the number of features or predictors. On one or more hidden layers is where the extraction of relevant features and the learning of complex patterns are performed. Finally, in the output layer, the observations are classified according to their class probability. In this case, the output layer has two neurons, because it is a binary classification task between CC and noCC.

MLPs have been widely used in several convective-hazard and precipitation forecasting studies (e.g., [58,59]); however, the application of Deep Neural Network architectures has been preponderant recently in this domain, i.e., Deep Neural Networks (DNN; e.g., [4,60,61], Recurrent Neural Networks (RNN; e.g., [62,63]), and CNN (e.g., [29,64,65]) are the most common). The implementation of Deep Learning techniques in deep convective cloud research will be addressed in later sections.

- Bagging and Random Forest

Bootstrap Aggregation, better known as the Bagging method, is a technique used to improve accuracy, reduce variance, and avoid overfitting by combining the predictions of multiple models trained on different subsets of the training data created through bootstrap sampling, where data points are randomly selected with replacements. (Figure 3a [66]). When sampling is performed without replacement, it is called pasting [67,68].

RF is a nonparametric ensemble model based on the consensus of a collection of DTs on different bootstrap samples. The outputs of the individual trees are then aggregated through voting to make the final prediction. Along with LR, RF is one of the most widely used ML models in issues related to convective-hazard forecasting, mostly because of its ability to capture non-linear association patterns between predictor and predictand, such as a convective storm system or precipitation [69]. Examples of RF applications in the literature are the works of [2–4,17,22,28,47,48,70–72].

In this study, two Bagging approaches were implemented, RF, and a Bagging method that used LR as the predictor in the bootstrap samples instead of DT (henceforth referred to as Bagging; Table 3 and Figure 3). The number of subsets of the training data corresponding to the number of estimators for RF was determined using a hyperparameter analysis for Los

Mochis and Mexico City datasets. In the case of Bagging, the default number of estimators (represented with the parameter n_estimators) in the scikit-learn library was set to 10.

- Voting

The voting method refers to an assembly strategy used to make predictions by integrating the collective predictions of multiple models or classifiers that are trained independently and in parallel on the same dataset. This technique aggregates the prediction of each classifier (here LR, DT, RF, SVM, and MLP; Figure 3b) and predicts the class that receives the most votes. This majority-vote classifier is called a hard voting classifier (Hvoting), otherwise, it is a soft-voting classifier (Svoting), where the class with the highest averaged class probability is selected as the final prediction [68]. Voting mitigates the impact of individual model biases and errors, leading to improved overall accuracy and robustness. For instance, the work of [3] found that majority voting effectively removed salt-and-pepper noise in its results for Convective Initiating objects predicted with DT and RF.

- Stacking

Also known as stacked generalization or stacked ensemble, this technique is used to combine predictions from multiple models to improve overall performance [73]. In this assembly strategy, the aggregate models called base models (LR, DT, RF, SVM, and MLP; Figure 3c) are used according to their weights to produce an output that is taken using a Meta-Classifier as the input. Each prediction from the base models becomes a new feature in the Meta-Classifier Dataset. The main purpose of the stacking method is that the Meta-classifier learns to combine the predictions of the base models optimally, leveraging their individual strengths. In this work, LR and RF were used as Meta-Classifiers, because both models reported the best performance metrics individually (LRstacking and RFstacking).

3.2.3. Machine Learning Process

In order to compare the performance between the different ML approaches, they were trained and tested with the same dataset configuration for each of the test sites. Before training and testing each of the ML approaches, a feature importance analysis or relative variable importance analysis was performed to estimate the degree of contribution of each of the interest fields in the prediction stage. In this context, RF and LR provide the relative importance of input variables when developing models such as attribute usage, Mean Decrease Impurity (MDI), and the absolute value of weighting coefficients, respectively [3]. In particular, feature importances of RF are provided by the fitted attribute feature_importances_ (of the scikit-learn python library). This approach evaluates the importance of each feature by measuring how much it contributes to reducing impurity, specifically Gini impurity, as part of the iterative process of feature-based splitting. The relative feature importance analysis helps to reduce dimensionality, improve model efficiency, and mitigate the risk of overfitting from the selection of predictor variables with the highest degree of contribution to the identification of the CC class. A lower dimensionality dataset improves model performance by removing noise from lower weight attributes, while computational costs are substantially reduced. In this sense, Los Mochis and Mexico City datasets were delimited by reducing to six the number of features used as predictor variables. For this task, the predictor selection criteria were the values with the greatest relative variable importance in LR and RF.

After selecting the predictors and scaling the data, hyperparameter tuning was performed using GridSearchCV, a function provided by the scikit-learn library that performs an exhaustive search over a specified parameter grid to find the best combination of hyperparameters for a machine learning model. In this work, the cross-validation strategy was used during the hyperparameter search which selected 5 k folds. The resulting optimal hyperparameter values are presented in Table 4. Finally, the performance of the models was evaluated using various accuracy metrics.

Table 4. The hyperparameters of LR, RF, and MLP were tuned using the cross-validation strategy.

Hyperparameter	Los Mochis	Mexico City
Norm of the penalty (LR)	L2	L2
The inverse of regularization strength (LR)	0.01	0.01
Solver (LR)	liblinear	newton-cg
Maximum depth of the tree (RF)	2	5
Number of trees in the forest (RF)	500	100
Random split predictor variables (RF)	1	1
Number of neurons in the hidden layer (MLP)	10	100
Solver (MLP)	sgd	Adam

3.2.4. Workflow and Implementation

A quasi-real-time modeling framework was designed by considering the following three principles: the use of open-source software and open-access products, automation, and scalability (Figure 4). Regarding the principle of using only free-access products, all processes and the workflow were implemented with Python 3.8 programming and were designed using ABI and GLM products from GOES R as inputs.

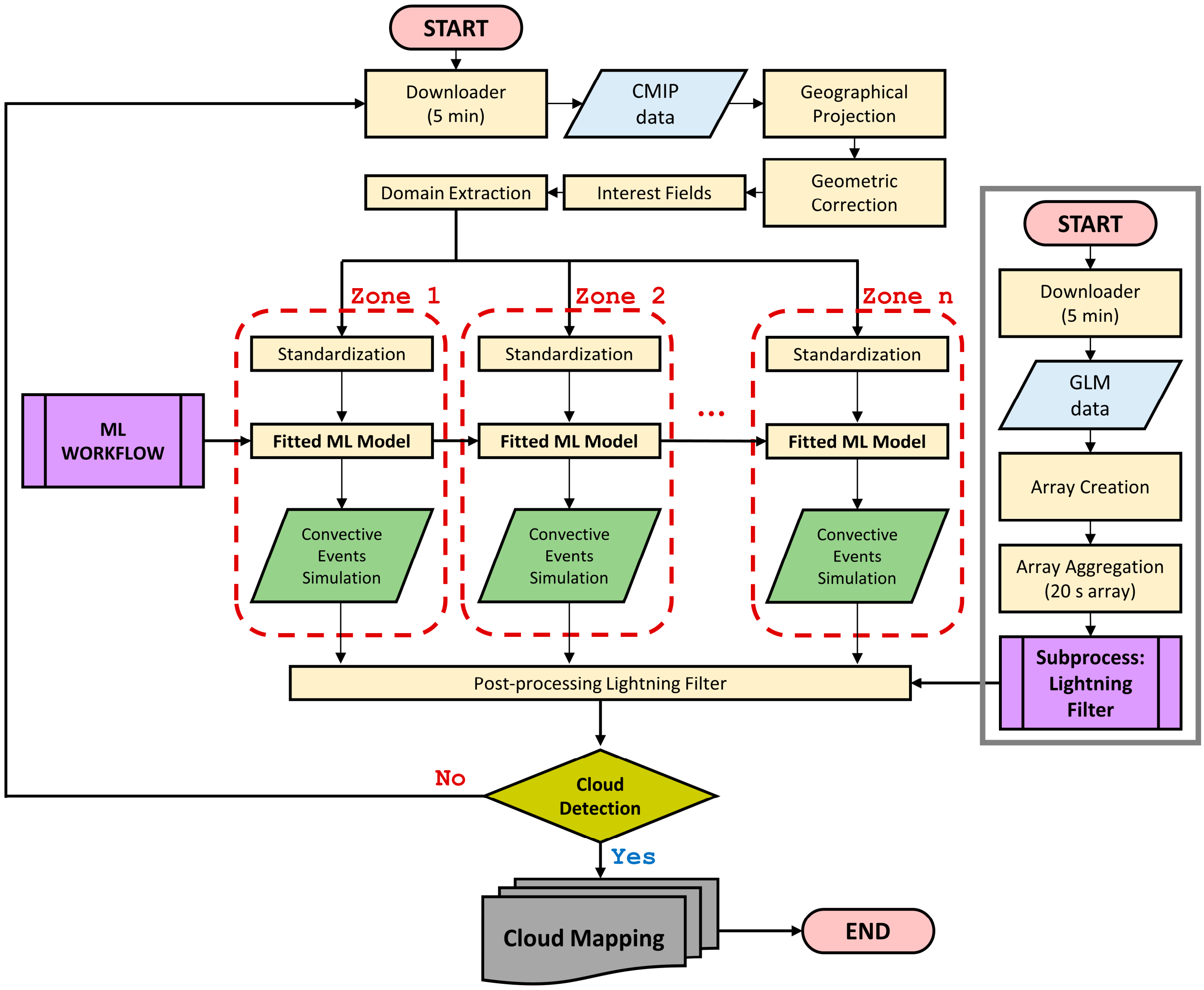


Figure 4. General scheme of the framework for identifying and monitoring potential convective events in quasi-real time (5 min). The ML approaches that are applied in the ML process are selected based on their performance parameters during the training and testing phases at each study site. CMIP: Cloud and Moisture Imagery.

The workflow runs cyclically and automatically every 5 min, corresponding to the temporal resolution of the ABI bands, and it begins with the download of ABI and GLM information from Amazon Web Service (AWS) S3 Storage (https://registry.opendata.aws/noaa-goes/, accessed on 5 December 2023) using Boto3, the AWS Software Development Kit (SDK) for Python. Concerning scalability, this novel methodology contemplates spatial extrapolation to the entire country of Mexico, following the principles of open access and automation, due to the availability of modeling inputs within the entire territory. However, for the training and testing phase of the ML models, zones with a homogeneous convective behavior must be delimited.

The spectral information that this framework is fed with (Table 1) is found in the GOES level 2 product, Cloud and Moisture Imagery (CMIP), which contains the information of the geostationary ABI bands in a NetCDF file [74]. Applying the parameters and attributes of the ABI instrument specified in the NetCDF file, projection to a mapping spatial domain is performed. Subsequently, the Interest Fields are calculated and spatial extraction of the study domains is performed.

In the second step, the pipelines (i.e., assembly of sequential data manipulation operations) of preprocessing and modeling are fed with the Interest Fields. Previously, at each site of interest, the models were trained and tested based on the creation of a set of reference historical events. When the optimal models and hyperparameters, as well as the significant variables that will be used as predictors, are identified, these configurations are stored in a pipeline (dashed red border Figure 4). The processing pipeline considers the scaling phase of the predictors from the StandardScaler method. The pipeline methods were taken from scikit-learn.

Pipeline outputs are mapped using the Python Matplotlib Basemap Toolkit library, which displays the geospatial information. A lightning incidence-based post-processing filter is additively integrated to identify potential deep convection zones not detected by ML models. The process runs iteratively again after 5 min.

3.3. Post-Processing Lighting Filter

To strengthen the output of the models, a post-processing lightning incidence-based filter has been integrated into the CC identification and monitoring framework. According to [2], the presence of lightning is a reliable indicator of intense convective activity typical of storm cloud formation environments. Several studies related to the detection of deep convective clouds include lightning data [3,22–24,64,75,76].

The GLM GOES data are used to generate a lightning matrix that is additively integrated into the output of the ML models to increase the probability of detection (POD) of the models. The GLM variable "Event" was selected, which represents the occurrence of a single-pixel exceeding the brightness detection threshold during one ~2 ms frame [77]. These products contain geospatial information on the incidence of lightning with a temporal resolution of 20 s. In this framework, the density or number of lightning strikes is not considered, and only the presence or absence of the meteor is taken as an indicator of CC. From the spatial location where the lightning is detected, an algorithm based on Euclidean distance (Equation (2)) is applied to associate it with the nearest cell center of the spatial domain of the Interest Field.

$$\mathrm{Distance}_{(\mathrm{LI,Ce})} = \sqrt{\left(\mathrm{Longitude}_{\mathrm{Ce}} - \mathrm{Longitude}_{\mathrm{LI}}\right)^2 + \left(\mathrm{Latitude}_{\mathrm{Ce}} - \mathrm{Latitude}_{\mathrm{LI}}\right)^2} \quad (2)$$

where C_e is the cell center of the Interest Field spatial domain and LI is the coordinate where the Lightning Event was registered. In addition, a 3 × 3 cell buffer was designated around the CC cell identified from the lightning filter (Figure 5).

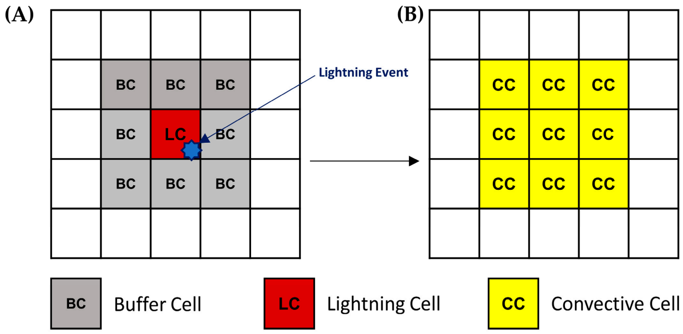


Figure 5. Schematization of the Lightning Filter generation process. When the presence of a Lightning Event is identified from the GOES GLM data accumulated in 5 min within the site of interest; it is associated with a cell of the model special domain, using its distance from the nearest cell center as the criterion assignment. The Lightning Cell (LC) together with a buffer of one cell per edge (BC) (**A**), are labeled as Convection Cells (CC) and are integrated into the output matrix (**B**) of the ML models.

3.4. Accuracy Metrics

To assess the performance of the different ML approaches, well-known classification metrics were calculated from the confusion matrix, including the Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), Total Classification Accuracy (Acc), Bias (BIAS), Precision, F1 Score (F1), and Intersection Over Union (IoU):

$$\mathrm{POD} = \frac{TP}{TP + FN} \tag{3}$$

$$\mathrm{FAR} = \frac{FP}{FP + TP} \tag{4}$$

$$\mathrm{CSI} = \frac{TP}{TP + FP + FN} \tag{5}$$

$$\mathrm{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\mathrm{BIAS} = \frac{TP + FP}{TP + FN} \tag{7}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\mathrm{F1} = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{POD}}{\mathrm{Precision} + \mathrm{POD}} \tag{9}$$

$$\mathrm{IoU} = \frac{TP}{TP + FP + FN} \tag{10}$$

where TP is the number of CC pixels that were correctly classified as CC (i.e., true positives), FP indicates the number of CC pixels that were incorrectly detected as CC (i.e., false positives), FN is the number of CC pixels that were incorrectly marked as noCC (i.e., false negative), and TN is all the remaining pixels that were correctly classified as noCC.

4. Results and Discussion

4.1. Selection of Interest Fields

Table 5 shows the Interest Fields selected according to their relative variable importance for CC cell detection with the LR and RF models for Los Mochis and Mexico City. In both study areas, according to the principal component analysis (PCA), the number of principal components that collectively explain 95% of the total variance in the datasets was five, which means that the dimensionality of the datasets is potentially reduced to five components while still retaining a significant amount of information from the original dataset. Here, the first five Interest Fields with the highest weighting coefficients in LR and highest MDI values in RF were selected. In both cases, the Interest Fields with the highest degree of contribution in the detection of CC cells coincided with four predictor variables, CtT, CtH01, CtH02, and CtH03. Accordingly, the dimensionality of the original datasets was reduced from 12 to 6 components.

Table 5. Interest Fields were used as predictor variables after the feature importance analysis, for Los Mochis and Mexico City datasets.

Los Mochis	**Mexico City**
CtT	CtT
CtH01	CtH01
CtH02	CtH02
CtH03	CtH03
CtG01	CtG01
CtG03	TChCtH03

The Interest Fields that recur in both models and zones are the most significant for the detection of deep convective clouds. For example, the Interest Field CtT, corresponding to the values of cloud-top temperature detected using the ABI-GOES 14 channel (11.2 µm TB), is one of those with the highest degree of contribution in both study sites, because the temperature at the top of the cloud is cooler than the surrounding environment being a typical property of deep convective clouds. During the convection process, this type of cloud reaches significant altitudes to regions where the atmospheric pressure is lower, giving way to the adiabatic cooling process, where the rising air expands and cools as it moves upward through the atmosphere [78], resulting in the temperature at the top of the cloud presenting significantly lower values. CtT is the Interest Field with a high degree of importance that is recurrent in other works; for example, Refs. [3,17,18] reported that CtT greatly contributed to the identification of Convective Initiation event (i.e., the probability that a given cumulus cloud object will develop into a $\geq$35-dBZ-intensity radar echo at -10 °C altitude [9]). In addition, in [28,79], CtT was identified as the most contributing variable to overshooting convective cloud tops' (OTs) classification using RF and LR models.

The following Interest Fields in order of importance are CtH01, CtH02, and CtH03. These spectral differences between ABI-GOES channels provide information on cloud-top height (cloud depth [3]). Similar to CtT, CtH01 is one of the variables with a higher degree of contribution in the detection of CC cells. This Interest Field is used to determine lower stratospheric moisture, and it is positive when water vapor is present above the cloud tops, which is an indicator of the presence of OTs [28]. For the glaciation indicators, only CtG01 was determined as one of the most significant variables.

Despite the time trend variables proposed as Interest Fields providing information on the rate of vertical cloud-top growth, in general, these variable predictors resulted in relatively lower contributions in both LR and RF models (except TChCtH03 in the Mexico City site).

The distribution of Interest Fields selected after feature importance analysis for Los Mochis and Mexico City are shown in Figures 6 and 7, respectively. CC and noCC sets are statistically differentiable in the six selected Interest Fields (which improves the classification task in the ML models) for the Los Mochis area. For the CC class, there are extremely small Inter Interquartile Ranges (IQRs) with a median close to 0 in the case of the spectral difference fields, and in that sense, ref. [28] claims that as a cloud reaches its local equilibrium level or the height of the tropopause, all channel differences will be almost zero. On the other hand, with the noCC class, the data distribution is much broader. In the case of the cloud-top temperature (CtT) field, the cells present higher values because the tops of the deep convection clouds are colder due to adiabatic cooling. In the case of the spectral difference fields, these mostly present negative values.

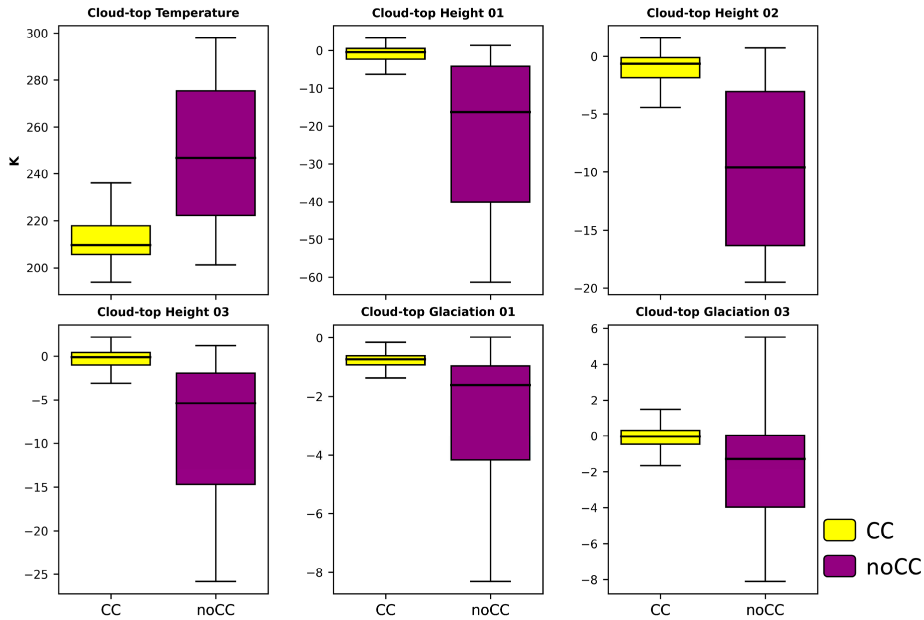


Figure 6. Box plots of input variables generated based on the reference data used for CC detection models, for Los Mochis. The predictor dataset was reduced from 12 input variables to 6 after analyzing the features in the LR and RF models. The number of classes labeled as CC is 24,116 and noCC is 10,884.

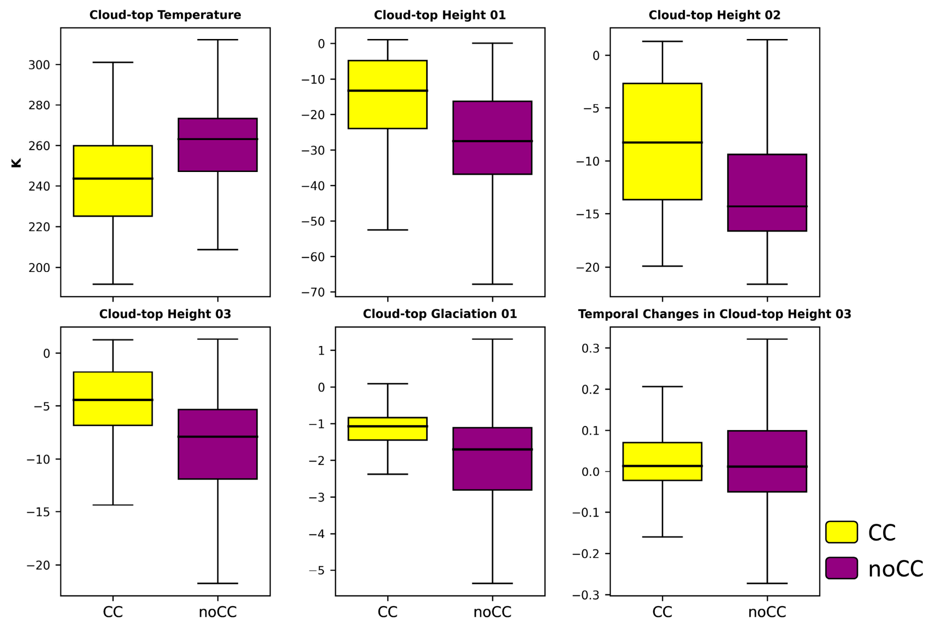


Figure 7. Box plots of input variables were generated based on the reference data used for CC detection models, for Mexico City. The predictor dataset was reduced from 12 input variables to 6 after analyzing the features in the LR and RF models. The number of classes labeled as CC is 57,833 and noCC is 79,667.

For the case study of Mexico City, overlapping distributions can be seen in the boxplots of the selected Interest Fields, which are reflected in the performance metrics of the ML models, which are notably lower than the metrics in the case of Los Mochis (Section 4.2). However, the general trends are preserved, and the CtT values are lower for the CC class; moreover, in the spectral difference fields of both height and glaciation, the noCC values are lower. Concerning the IQRs of the predictors, the distributions are more balanced for both classes, which, together with the distribution overlap, allows us to infer greater complexity in the modeling of CC events than at the Los Mochis site. Returning to the CtT Interest Field, its high degree of contribution to the detection of the CC class has been clearly shown, and cells with lower temperature values are associated with the presence of potential storm clouds. In this context, ref. [80] lists some cloud-top temperature thresholds reported in studies related to the identification of deep convection clouds, for example, 241 and 221 °K [81], 221 °K [82], 225 °K [83], 255 and 206 °K [84]. Here, it was estimated that storm clouds manifest at a CtT threshold of <220 °K for the Los Mochis study site and <260 °K for Mexico City.

4.2. Performance and Validation of Detection Models

Each ML approach was evaluated on the basis of six performance metrics that indicate each model's ability to detect convective cells, false alarm rates, etc. The effect of a post-processing filter based on lightning incidence was also evaluated. Figures 8 and 9 show the results for datasets corresponding to Los Mochis and Mexico City, respectively.

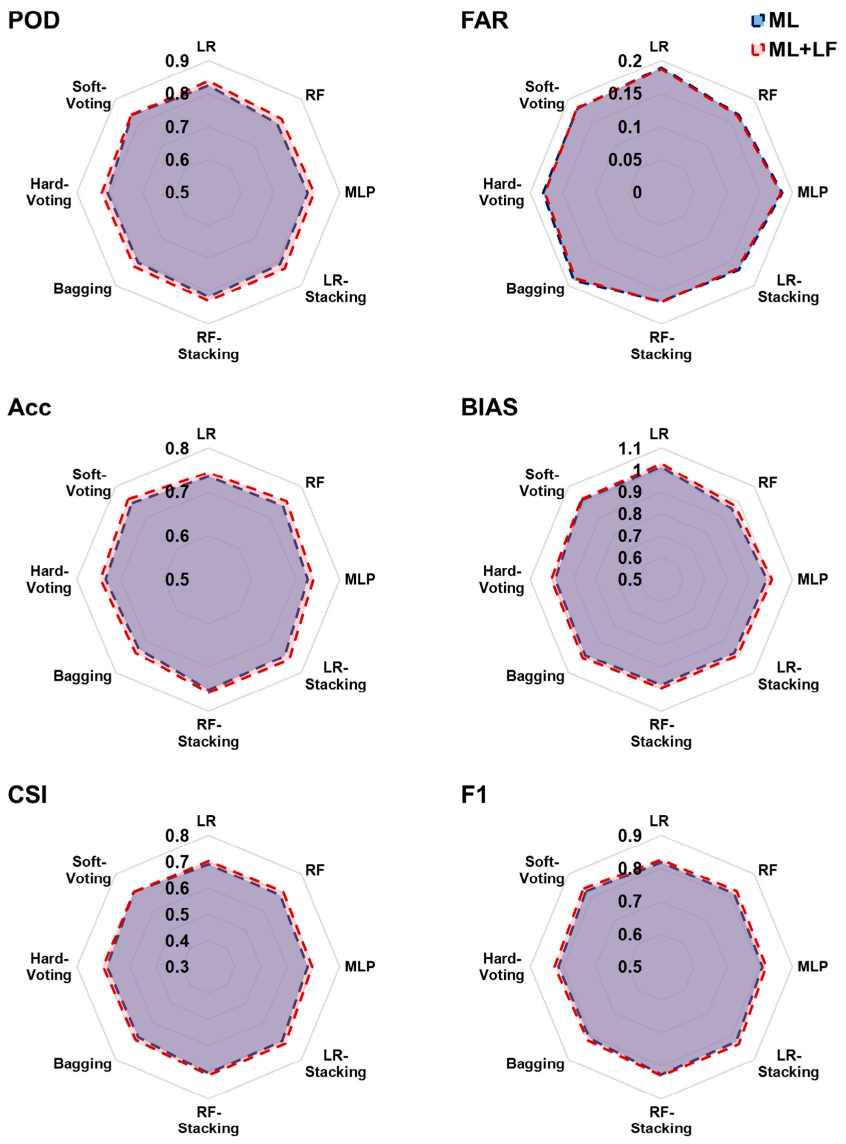


Figure 8. Assessment results of POD, FAR, Acc, BIAS, CSI, and F1 metrics for the ML approaches using Los Mochis test dataset. Each vertex of the graph corresponds to an ML approach, the blue lines correspond to the performance results of the ML models (ML), and the red lines show the results integrating the Lightning Filter (ML + LF).

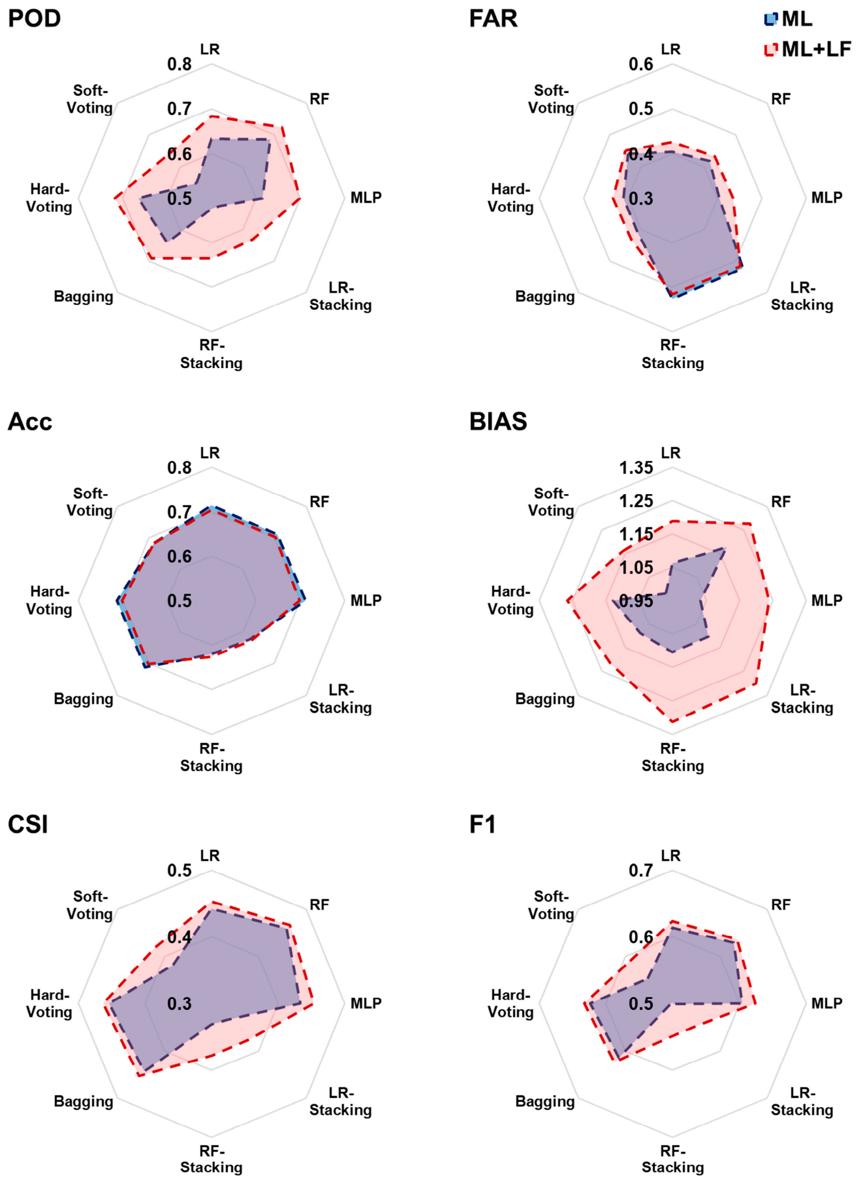


Figure 9. Assessment results of POD, FAR, Acc BIAS, CSI, and F1 metrics for the ML approaches using the test dataset in Mexico City. Each vertex of the graph corresponds to an ML approach, the blue lines correspond to the performance results of the ML models (ML), and the red lines show the results integrating the Lightning Filter (ML + LF).

For the Los Mochis site, there is a small difference between the results of the ML approach (ML series) and those with the integrated post-processing filter (ML + LF). In this regard, the ability of the models to detect the CC class, represented by the POD metric, presents values close to 0.8, while the false alarm ratio (FAR) in all the ML approaches is lower than 0.2. This means that the presence of potential deep convection events can be efficiently detected without the need to integrate the lightning incidence data. However, this does not indicate that the presence of lightning is uncommon in this zone; on the contrary, Ref. [85] reported a significant lightning strike density in northeastern Mexico. The region with the highest presence of lightning activity is in the narrow strip between the Sierra Madre Occidental and the Gulf of California, with a maximum in the months of July and August, corresponding to the dates used to construct the datasets of this research work. In addition, the presence of lightning strikes from the GLM-GOES was detected in 8 of the 14 images used to construct the training and test datasets.

Nevertheless, the Mexico City site shows significant changes when the Lightning Filter is integrated, especially favoring the POD, CSI, and F1 metrics. For example, the highest value was POD = 0.68, obtained using RF, which increased to POD = 0.72 after post-processing. The second-highest POD was achieved with LR, which increased from POD = 0.63 to POD = 0.7 with the integration of the Lightning Filter. Comparatively to the Los Mochis dataset, in Mexico City the integration of GLM data significantly improves the system's ability to detect the CC class in all ML approaches because lightning strike incidence is a reliable indicator of a deep convection event generation, and, like the Los Mochis site, this zone has intense lightning activity. In addition to POD, the CSI and F1 metrics increased in ML approaches with lower overall performances, such as LRstacking and Soft-Voting. On the other hand, LF has a negative impact on the BIAS metric, increasing the values above the optimum of 1. This metric ranges from 0 to infinity and allows us to assess whether the forecasting method tends to underestimate the CC (BIAS < 1) or overestimate it (BIAS > 1 [86]). In this sense, this metric, by presenting values greater than 1 when integrating the Lightning Filter, could reveal one of the limitations of labeling convective cells with the MODIS sensor, which has a lower resolution than the ABI-GOES products, since there are areas labeled as noCC that present lightning activity that clearly indicates the generation of a convective environment. Therefore, in future work, it will be proposed to include unsupervised learning approaches that simplify the task of generating a reference set with another sensor.

Overall, for the Los Mochis dataset, no significant differences were found between the use of model ensembles and simpler approaches such as LR, which is consistent with the findings of [17,23,28], who reported that this model is a robust alternative for convective-hazard modeling and detection. In contrast, for the Mexico City dataset, significant differences in the performance of each ML approach were observed. These differences are variable for each metric, but it is perceivable that LRstacking, RFstacking, and Soft-Voting are the ML approaches with the poorest performances. Instead, for all six metrics, LR, RF, MLP, Bagging, and Hard-Voting showed similar performance.

Regarding the POD values, ultimately, the acceptable level of POD may vary depending on the specific goals and requirements of the forecasting or detection system. However, a high POD may be necessary to minimize the risk of missing hazardous events. In this sense, for the Los Mochis dataset, the highest value was POD = 0.84, while in the Mexico City dataset it was POD = 0.72, with both estimated with LR after post-processing filtering. In this context, Refs. [28] and [29] report performances of POD = 0.77 and POD = 0.79, respectively, in their work on Ots' detection, while [3] and [2], who studied CI detection, obtained values of POD $\approx$ 0.8 and POD > 0.7. Other examples are [87] with their proposed pre-convective-environment alerting and monitoring system, which obtained a POD between 0.66 and 0.7; on the other hand, [88] with his convective storm nowcasting based on CNN, POD values were close to 0.7.

The FAR metric provides information about how often the classifier makes incorrect positive predictions when it should not. In the case of Los Mochis, the FAR values were below 0.2, and these indices did not show significant variations among the different ML approaches, with values between 0.16 and 0.18. In contrast, in Mexico City, significant variations were observed, with the highest values being LRstacking and RFstacking with FAR = 0.51. For the rest of the ML approaches, it can be seen that LF increased the number of false alarms, but to a lesser extent. For example, for LR, a change from FAR = 0.40 to 0.42 was estimated, whereas for RF, the change was from FAR = 0.41 to 0.43. In the literature, FAR values vary significantly depending on the type of convective forecasting performed. For example, [28] and [29] reported FAR values = 0.3 and 0.09, whereas [89] obtained FAR values between 0.01 and 0.18. For CI detection, ref. [2] estimated FAR values between 0.46 and 0.83; [3] a FAR $\approx$ 0.2; in [17] the FAR ranges between 0.22 and 0.36; and in [18] 0.46 to 0.37.

Accuracy provides an overview of how effective a model is at correctly classifying samples. In the Los Mochis dataset, there are no significant differences between the Acc values of the ML approaches, which generally have values slightly higher than 0.7. In this case, the addition of LF increases the Acc value, but not substantially. In Mexico City, the lowest values were recorded for LRstacking and RFstacking with Acc = 0.62, while LR, RF, Bagging, and Hard-Voting had Acc values of 0.7. This metric is susceptible to bias in datasets with class imbalance because it does not consider the distribution of classes in the dataset. Therefore, it is important to complement the performance analyses of the models with other performance metrics. Here, the Acc values between Los Mochis and Mexico City were more similar than the other metrics because there was more class imbalance in the Los Mochis (CC class = 69%, noCC class = 31%) than in the Mexico City dataset (CC class = 42%, noCC class = 58%).

BIAS is another common metric in studies related to the detection of convective events (e.g., [86]), as well as in the forecasting of convective hazards such as lightning strikes (e.g., [24]) because it evaluates whether the forecast method tends to underestimate or overestimate the CC occurrence. The estimated BIAS values for Los Mochis range from 0.97 to 1.03, showing a relatively good performance in terms of false alarms and misses. In contrast, in Mexico City, there is an important variation between the ML approaches, with values up to BIAS > 1.15 for RF, which increases to BIAS > 1.25 after LF. The analysis of different metrics allows for the evaluation of the forecasting framework proposed in this study. In this sense, the integration of data from the GLM increases the overestimation rate of the CC class concerning the reference target dataset, but it favors the ability of the system to detect the occurrence of CC.

CSI measures the accuracy of a forecast in predicting the occurrence of a specific event (e.g., severe thunderstorms) relative to observations. It is particularly useful in situations where false alarms or missed events can have significant consequences, such as severe weather forecasting. For Los Mochis, the dataset values close to 0.7 have been found; therefore, a CSI value greater than 0.7 indicates a good classification model. This means that the model is reasonably accurate in identifying positive cases while minimizing false alarms and misses. The CSI results for the Mexico City dataset range between 0.4 and 0.5, with a slight increase after LF, which is higher for the worst-performing metrics such as LRstacking, RFstacking, and Soft-Voting. A CSI between 0.4 and 0.7 indicates a moderate level of success in the classification task. Although the model makes correct positive predictions, there is scope for enhancement in reducing false alarms.

Regarding the F1 metric, a high F1 score (closer to 1) indicates that the model has high levels of precision and recall. In other words, it correctly classifies positive instances while avoiding false positives and false negatives. From the performance of Los Mochis, it can be concluded that all ML approaches can correctly classify positive cases, controlling false positives and false negatives. Similar to the CSI metric, in the Mexico City dataset, the post-processing LF improves the F1 score of the models with the lowest performance, whereas in the ML approaches, LR, RF, MLP, Bagging, and Hard-Voting, F1 values are estimated at F1 = 0.63.

Figure 10 presents the Receiver Operating Characteristic (ROC) curves for the ML approaches for the Los Mochis and Mexico City datasets. The ROC curves demonstrate the trade-off between sensitivity and specificity at different threshold values. A classifier with a curve closer to the upper-left corner indicates superior performance because it achieves high true-positive rates while maintaining low false-positive rates. For Los Mochis, homogeneous behavior was observed among the ML approaches, except for RFstacking, which showed a lower discrimination capacity. The area under the ROC curve (AUC) values is higher than 0.7, which means that, in general, the ML approaches have an acceptable discrimination capacity between classes. In the case of Mexico City, the ROC curve analysis revealed that LR, RF, MLP, Bagging, and Hard-Voting consistently outperformed the other classifiers across various threshold values. Their ROC curves exhibit a steeper ascent, indicating better discrimination between CC and noCC cases. This result suggests that these ML approaches are well suited for CC detection in this particular dataset. However, it is essential to consider the information provided by the other metrics.

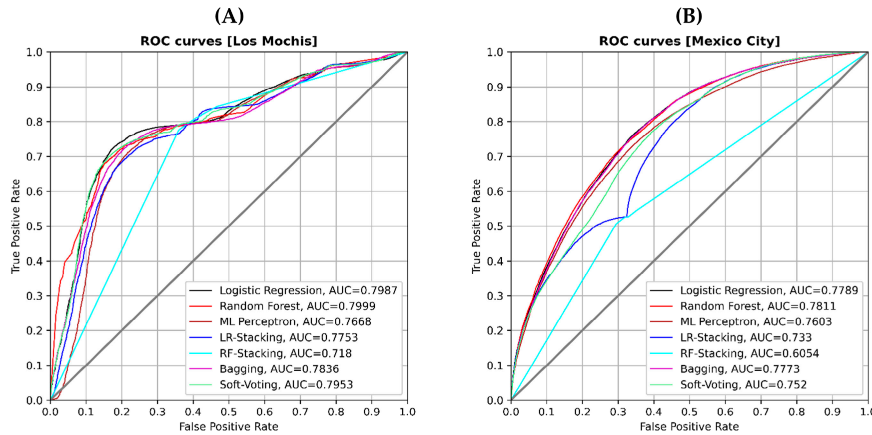


Figure 10. ROC curves and AUC values for the predictions obtained with LR (black), RF (red), MLP (brown), LRstacking (blue), RFstacking (cyan), Bagging (magenta), and Svoting (light green), for Los Mochis (**A**) and Mexico City (**B**). These ROC curves provide a comparison of the sensitivity and specificity of each ML approach, at different discrimination thresholds.

Examples of a deep convection event simulated from each ML approach for Los Mochis and Mexico City are shown in Figures 11 and 12, respectively; this comparison allows for a qualitative evaluation of the performance of each model, while the IoU metric indicates the overlap between the predicted regions and the ground truth (Table 6).

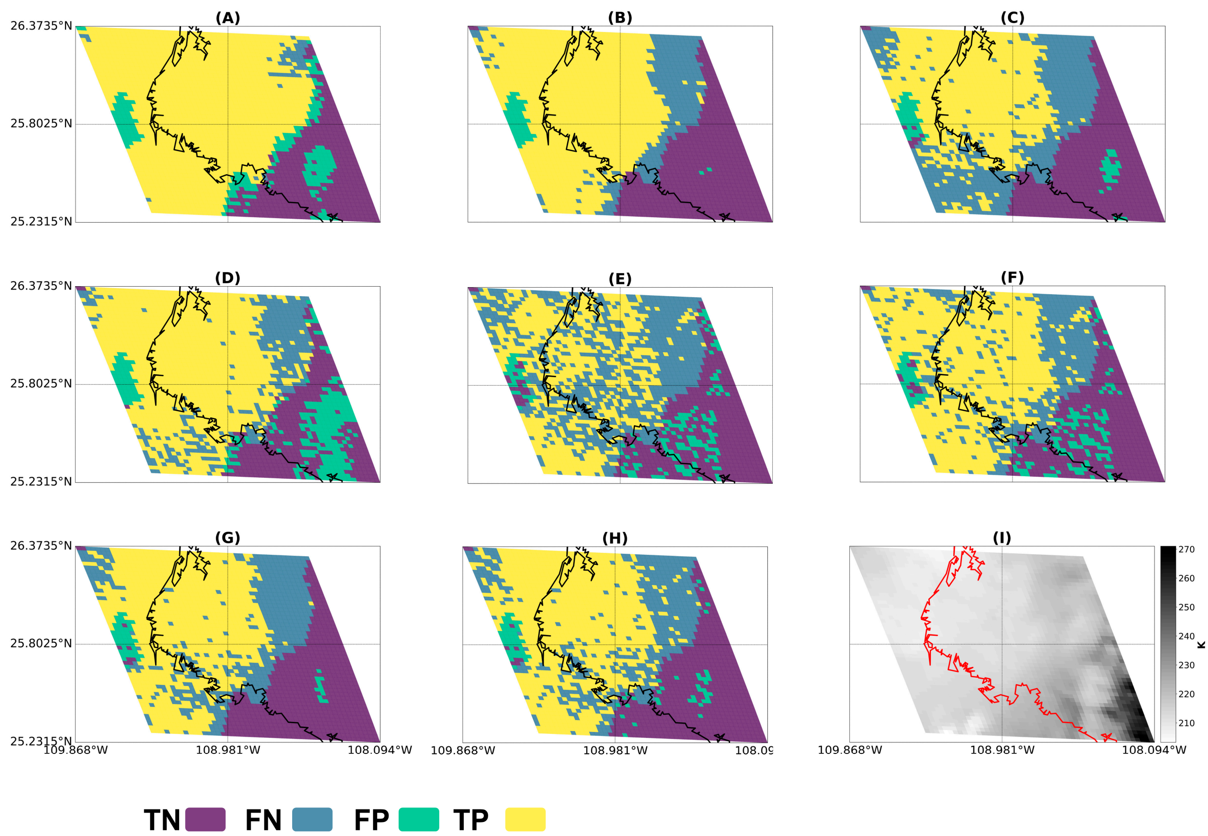


Figure 11. IoU maps of the Los Mochis 2018-227—18:00 event (15 August 2018) generated with the reference labels and simulations of the eight ML approaches LR (**A**), RF (**B**), MLP (**C**), Bagging (**D**), LRstacking (**E**), RFstacking (**F**), Hvoting (**G**), Svoting (**H**), and cloud-top brightness temperature map of the same event obtained from the ABI Band 14 (**I**). In general, it is observed that all the models underestimate the edge of the cloud, except LR, which presented values IoU = 0.86 and POD = 0.8.

Table 6. IoU metric results for the Los Mochis 2018-227—18:00 and Mexico City 2019-247—20:10 (4 September 2019) events for each ML Approach.

Event	LR	RF	MLP	LRstacking	RFstacking	Bagging	Hvoting	Svoting
Los Mochis 2018-227—18:00	0.86	0.76	0.56	0.44	0.62	0.67	0.65	0.65
Mexico City 2019-247—20:10	0.53	0.51	0.51	0.41	0.42	0.53	0.52	0.45

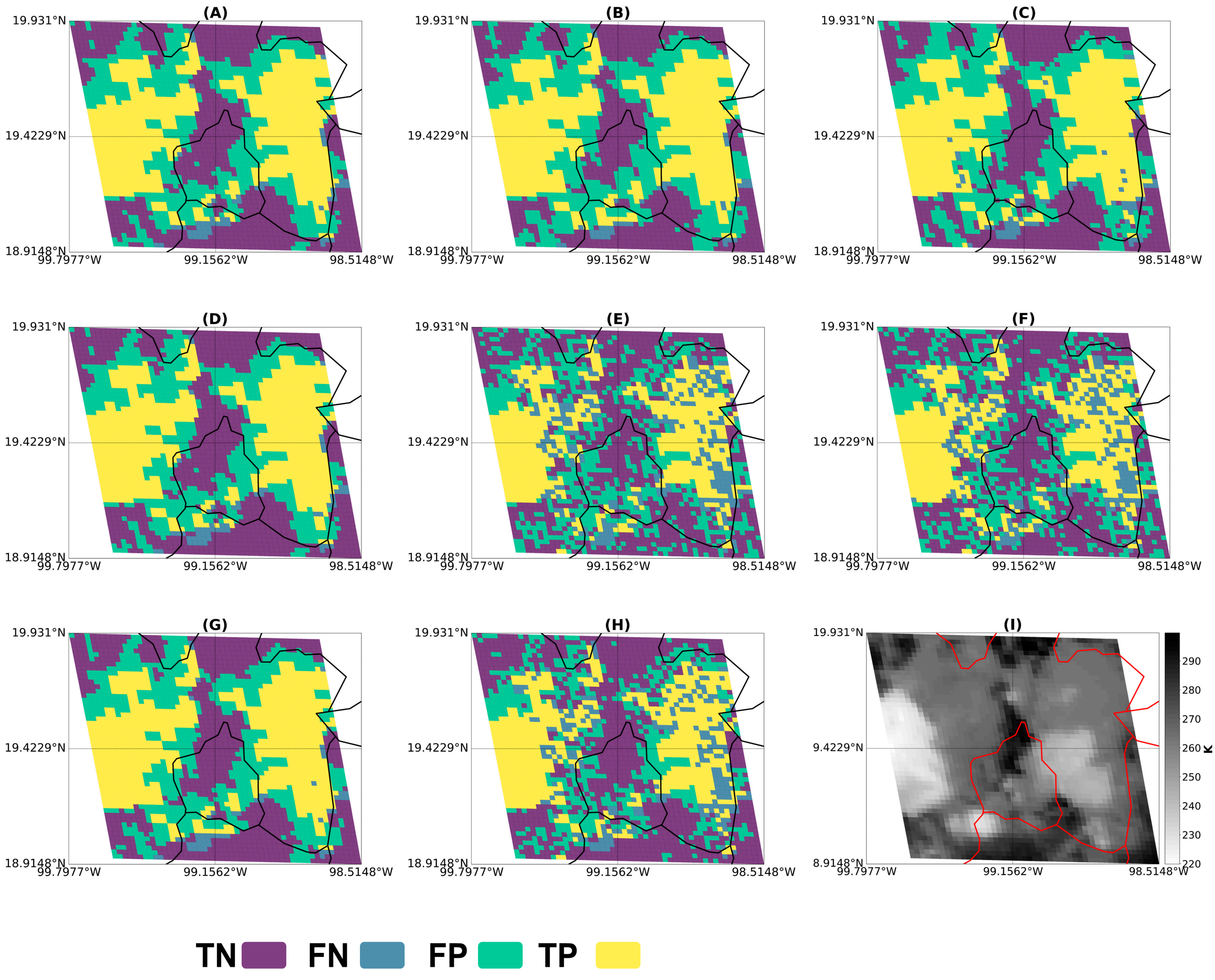


Figure 12. IoU maps of the Mexico City 2019-247—20:10 event generated with the reference labels and simulations of the eight ML approaches LR (**A**), RF (**B**), MLP (**C**), Bagging (**D**), LRstacking (**E**), RFstacking (**F**), Hvoting (**G**), Svoting (**H**), and cloud-top brightness temperature map of the same event obtained from the ABI Band 14 (**I**). The outputs of LR, RF, MLP, Bagging, and Hvoting are consistent with each other, showing a significant overestimation at the edge of the CCs. On the other hand, the outputs generated with LRstacking, RFstacking, and Svoting are extremely noisy, especially at the boundary of the clouds.

In the case of Los Mochis, the simulation of the event that occurred on 15 August 2018 shows a tendency to underestimate the total area of the deep convective cloud at the cloud edges, with the exception of LR where the IoU = 0.86 values are the highest recorded. The presence of false-negative pixels at the cloud edge indicates the difficulty of simulating this transition zone where cloud properties estimated from ABI-GOES data become diffuse. Therefore, future work will consider adding an additional transition class between CC and non-CC. In most ML approaches, except RF, a convective core is simulated in the lower-right part of the simulation domain that does not match the reference labels for this event, which could be related to a better ability to detect deep convection events from AGI-GOES data than from MODIS.

The ensemble models LRstacking and RFstacking show the worst performance, with values IoU = 0.44 and IoU = 0.62, respectively. In these ML approaches, a pronounced salt-and-pepper effect is observed, which usually occurs in pixel-based classification tasks when no contextual spatial information is provided [90]. This pattern is also present in the simulated event for Mexico City (Figure 12), where LRstacking, RFstacking, and Svoting also show this effect, which is related to the low-performance metrics of the test dataset. In this event, two well-separated clouds can be seen forming over the areas of higher topographic elevation, which, in addition to the presence of nuclei with lower cloud-top temperature, allows us to infer that its formation was due to a forced convection process. However, as in Los Mochis, all models show difficulties in simulating the edge of the clouds, with the difference being that, in this case, the areas of both convective clouds are overestimated with the presence of false-positive pixels.

As in the previous case, this area of opportunity can be addressed by adding an additional transition class or using alternative approaches that maintain the spatial structure, such as Convolutional Neural Networks (CNN) or other spatially aware models.

4.3. Computational Costs of ML Approaches

Figure 13 presents a comprehensive evaluation of the computational costs associated with the eight ML approaches in Los Mochis (Figure 13a) and Mexico City (Figure 13b). The assessment considers both computing time and memory usage to provide insights into the efficiency and resource requirements of each model in a standardized computing environment to ensure consistency.

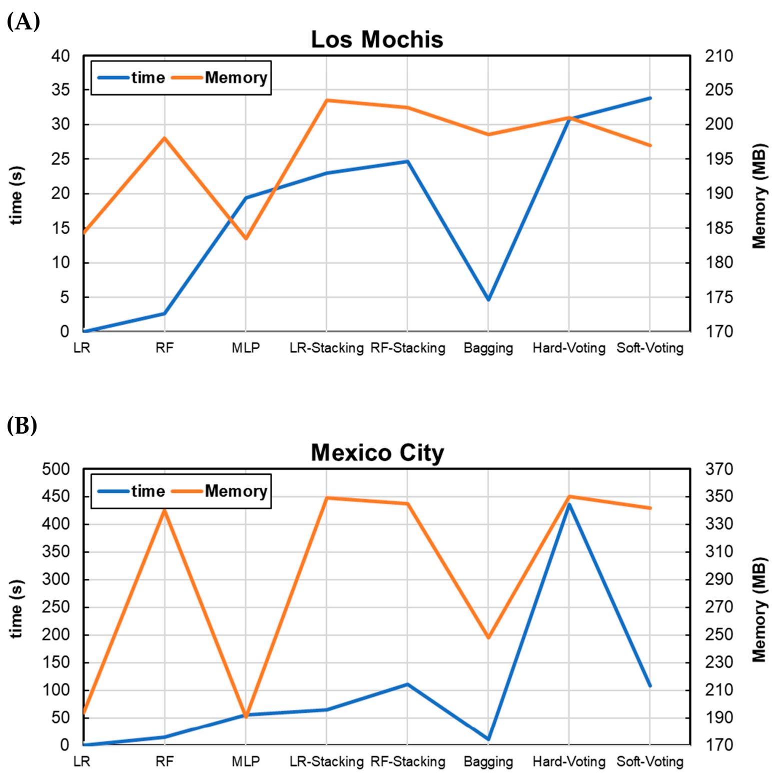


Figure 13. Computational costs chart for each ML approach corresponding to Los Mochis (**A**) and Mexico City (**B**). The right axis indicates the total simulation time, and the left shows the amount of memory usage.

For Los Mochis, the results revealed notable variations in computational costs across the evaluated ML approaches. RF and Bagging demonstrated efficient computing times but consumed higher memory, emphasizing their resource-intensive nature. LR and MLP presented balanced trade-offs between time and memory. In this case, in particular, the ensemble models imply a higher memory usage due to their higher degree of complexity, but this trend does not hold in the computational time section, where LR, RF, and Bagging were the ML approaches with less time consumption, and with a low significant variation.

In Mexico City, Hard-Voting showed a longer computation time compared to the other ML approaches; it is important to add that the omission of SVM in Soft-Voting reduces the computation time compared to Hard-Voting. In Los Mochis, RF showed efficient computation times but consumed more memory, while, LR, RF, and Bagging showcased competitive performance in terms of both time and memory usage.

A more comprehensive comparison should align with the specific requirements of the application, dataset size, and available computational resources.

5. Conclusions

In this study, a quasi-real-time (approximately every 5 min) framework for deep convective events modeling, based on ABI-GOES data, was designed by considering the following three principles: the use of open-source software and open-access products, automation, and scalability. This modeling framework runs cyclically and automatically every 5 min, corresponding to the temporal resolution of the ABI bands.

This framework is based on the use of machine learning techniques, for which eight different models and model ensemble strategies, including LR, RF, MLP, LRstacking, RFstacking, Bagging, Hard-Voting, and Soft-Voting, were compared for two study sites, Los Mochis and Mexico City, selected for their intense convective activity and high degree of vulnerability to extreme weather events. The results indicate that a simple approach such as LR or RF can be a good alternative for the identification and simulation of deep convective clouds in both study areas, showing POD $\approx$ 0.84 for Los Mochis and POD $\approx$ 0.72, while FAR $\approx$ 0.2 and FAR $\approx$ 0.4 values were estimated, respectively.

In addition, the implementation of a post-processing filter based on lightning incidence recorded by GLM-GOES was evaluated to improve the detection capability of the ML models. In this sense, it was observed that, in the case of Los Mochis, these did not represent a significant change in contrast to Mexico City, where the detection capability of all ML approaches increased significantly.

From the simulation of a reference event in both study areas, a challenge is identified in the simulation of cloud edges due to the limitations of using a pixel-based classification approach; however, the results show a good overall performance in most of the ML approaches analyzed.

This work sets a precedent towards the implementation of an early warning system for hazards associated with intense convective activity in a region with complex atmospheric dynamics such as Mexico, which also has a limited severe-weather-monitoring infrastructure and a highly heterogeneous distribution of ground-based sensors.

Author Contributions: Conceptualization, R.D.-O. and J.T.-V.; methodology, R.D.-O.; software, R.D.-O.; validation, J.T.-V.; formal analysis, R.D.-O., J.N.C.-P., J.A.V.-Z. and J.T.-V.; investigation, R.D.-O.; writing—original draft preparation, R.D.-O.; writing—review and editing, J.N.C.-P., J.A.V.-Z. and J.T.-V.; visualization, R.D.-O.; supervision, J.N.C.-P., J.A.V.-Z. and J.T.-V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Acknowledgments: This work was supported by the Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), through grant CVU: 841667. We thank the División de Geociencias Aplicadas of the Instituto Potosino de Investigación Científica y Tecnológica A.C. and the Geomatics and Numerical Modeling Laboratory for their personal and technical support. The authors express their sincere gratitude to the National Centers for Environmental Information (NCEI) at the NOAA for providing access to the GOES-R data.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABI	Advanced Baseline Imager
ANN	Artificial Neural Networks
AWS	Amazon Web Services
CC	Convective Cells
CI	Convective Initiation
CMIP	Cloud and Moisture Imagery
CNN	Convolutional Neural Networks
COMS	Communication, Ocean, and Meteorological Satellite
COT	Cloud Optical Thickness
CSI	Critical Success Index
CtP	Cloud-top Pressure
DL	Deep Learning
DT	Decision Tree
ENSO	El Niño–Southern Oscillation
FAR	False alarm ratio
GLM	Geostationary Lightning Mapper
GOES	Geostationary Operational Environmental Satellite
IoU	Intersection over Union
ISCCP	International Satellite Cloud Climatology Project
LF	Lightning Filter
LR	Logistic Regression
MDI	Mean Decrease Impurity
ML	Machine Learning
MLP	Multi-layer Perceptron
MODIS	Moderate-Resolution Imaging Spectroradiometer
NAM	North American monsoon
NWP	Numerical weather predictions
OT	Overshooting Top
POD	Probability of detection
RF	Random Forest
RNN	Recurrent Neural Networks
ROC	Receiver Operating Characteristic
SMO	Sierra Madre Occidental
SVM	Support Vector Machine

References

1. Cancelada, M.; Salio, P.; Vila, D.; Nesbitt, S.W.; Vidal, L. Backward Adaptive Brightness Temperature Threshold Technique (BAB3T): A Methodology to Determine Extreme Convective Initiation Regions Using Satellite Infrared Imagery. *Remote Sens.* **2020**, *12*, 337. [CrossRef]
2. Han, H.; Lee, S.; Im, J.; Kim, M.; Lee, M.-I.; Ahn, M.H.; Chung, S.-R. Detection of Convective Initiation Using Meteorological Imager Onboard Communication, Ocean, and Meteorological Satellite Based on Machine Learning Approaches. *Remote Sens.* **2015**, *7*, 9184–9204. [CrossRef]
3. Lee, S.; Han, H.; Im, J.; Jang, E.; Lee, M.-I. Detection of Deterministic and Probabilistic Convection Initiation Using Himawari-8 Advanced Himawari Imager Data. *Atmos. Meas. Tech.* **2017**, *10*, 1859–1874. [CrossRef]
4. Liu, Q.; Li, Y.; Yu, M.; Chiu, L.S.; Hao, X.; Duffy, D.Q.; Yang, C. Daytime Rainy Cloud Detection and Convective Precipitation Delineation Based on a Deep Neural Network Method Using GOES-16 ABI Images. *Remote Sens.* **2019**, *11*, 2555. [CrossRef]
5. León-Cruz, J.F.; Caetano, E.; Cortés-Ramos, J.; Dominguez, C.; Méndez-Pérez, J.M. Thunderstorm and Hailstorm Environments in Mexico. *Int. J. Climatol.* **2023**, *43*, 4379–4395. [CrossRef]
6. Weckwerth, T.M.; Parsons, D.B. A Review of Convection Initiation and Motivation for IHOP_2002. *Mon. Weather Rev.* **2006**, *134*, 5–22. [CrossRef]

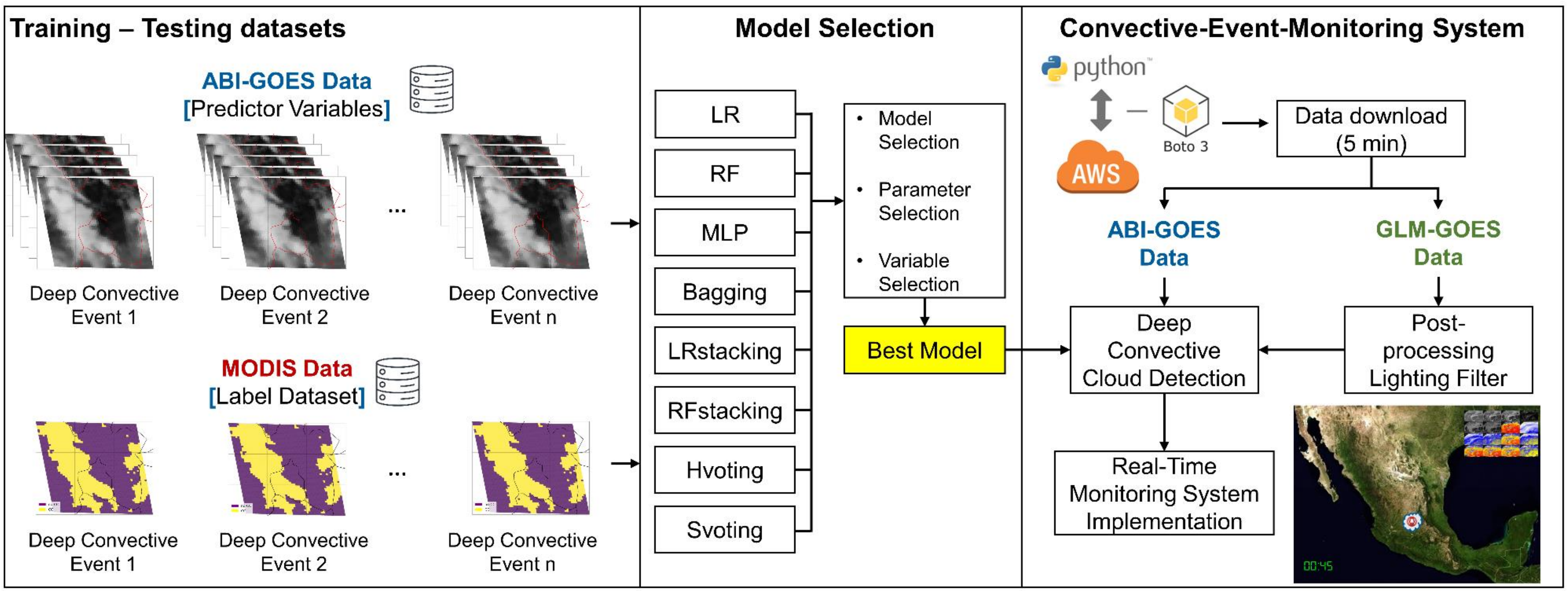
7. Dixon, M.; Wiener, G. TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-Based Methodology. *J. Atmos. Ocean. Technol.* **1993**, *10*, 785–797. [CrossRef]
8. Cintineo, J.L.; Pavolonis, M.J.; Sieglaff, J.M.; Wimmers, A.; Brunner, J.; Bellon, W. A Deep-Learning Model for Automated Detection of Intense Midlatitude Convection Using Geostationary Satellite Images. *Weather Forecast.* **2020**, *35*, 2567–2588. [CrossRef]
9. Gravelle, C.M.; Mecikalski, J.R.; Line, W.E.; Bedka, K.M.; Petersen, R.A.; Sieglaff, J.M.; Stano, G.T.; Goodman, S.J. Demonstration of a GOES-R Satellite Convective Toolkit to "Bridge the Gap" between Severe Weather Watches and Warnings: An Example from the 20 May 2013 Moore, Oklahoma, Tornado Outbreak. *Bull. Am. Meteorol. Soc.* **2016**, *97*, 69–84. [CrossRef]
10. Roberts, R.D.; Rutledge, S. Nowcasting Storm Initiation and Growth Using GOES-8 and WSR-88D Data. *Weather Forecast.* **2003**, *18*, 562–584. [CrossRef]
11. Mueller, C.K.; Wilson, J.W.; Crook, N.A. The Utility of Sounding and Mesonet Data to Nowcast Thunderstorm Initiation. *Weather Forecast.* **1993**, *8*, 132–146. [CrossRef]
12. Veillette, M.S. Convective Initiation Forecasts Through the Use of Machine Learning Methods. In Proceedings of the 93rd American Meteorological Society Annual Meeting, Austin, TX, USA, 9 January 2013.
13. Mecikalski, J.R.; Bedka, K.M. Forecasting Convective Initiation by Monitoring the Evolution of Moving Cumulus in Daytime GOES Imagery. *Mon. Weather Rev.* **2006**, *134*, 49–78. [CrossRef]
14. McGovern, A.; Chase, R.J.; Flora, M.; Gagne, D.J.; Lagerquist, R.; Potvin, C.K.; Snook, N.; Loken, E. A Review of Machine Learning for Convective Weather. *Artif. Intell. Earth Syst.* **2023**, *2*, e220077. [CrossRef]
15. Walker, J.R.; MacKenzie, W.M.; Mecikalski, J.R.; Jewett, C.P. An Enhanced Geostationary Satellite–Based Convective Initiation Algorithm for 0–2-h Nowcasting with Object Tracking. *J. Appl. Meteorol. Climatol.* **2012**, *51*, 1931–1949. [CrossRef]
16. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Prabhat Deep Learning and Process Understanding for Data-Driven Earth System Science. *Nature* **2019**, *566*, 195–204. [CrossRef] [PubMed]
17. Mecikalski, J.R.; Williams, J.K.; Jewett, C.P.; Ahijevych, D.; LeRoy, A.; Walker, J.R. Probabilistic 0–1-h Convective Initiation Nowcasts That Combine Geostationary Satellite Observations and Numerical Weather Prediction Model Data. *J. Appl. Meteorol. Climatol.* **2015**, *54*, 1039–1059. [CrossRef]
18. Han, D.; Lee, J.; Im, J.; Sim, S.; Lee, S.; Han, H. A Novel Framework of Detecting Convective Initiation Combining Automated Sampling, Machine Learning, and Repeated Model Tuning from Geostationary Satellite Data. *Remote Sens.* **2019**, *11*, 1454. [CrossRef]
19. Krinitskiy, M.; Sprygin, A.; Elizarov, S.; Narizhnaya, A.; Shikhov, A.; Chernokulsky, A. Towards the Accurate Automatic Detection of Mesoscale Convective Systems in Remote Sensing Data: From Data Mining to Deep Learning Models and Their Applications. *Remote Sens.* **2023**, *15*, 3493. [CrossRef]
20. Petersen, R.; Aune, R.; Rink, T. Objective Short-Range Forecasts of the Pre-Convective Environment Using SEVIRI Data. In Proceedings of the 2010 EUMETSAT Meteorological Satellite Conference, EUMETSAT, Cordoba, Spain, 20–24 September 2010.
21. Lenk, S.; Senf, F.; Deneke, H.; Hungershoefer, K. On the Sensitivity of Nowcasting Algorithms for Convective Initiation in Satellite Images to Automatic Tracking Techniques. In Proceedings of the 2018 EUMETSAT Meteorological Satellite Conference, Tallinn, Estonia, 17–21 September 2018.
22. La Fata, A.; Amato, F.; Bernardi, M.; D'Andrea, M.; Procopio, R.; Fiori, E. Cloud-to-Ground Lightning Nowcasting Using Machine Learning. In Proceedings of the 2021 35th International Conference on Lightning Protection (ICLP) and XVI International Symposium on Lightning Protection (SIPDA), Colombo, Sri Lanka, 20–26 September 2021; Volume 1, pp. 1–6.
23. Ukkonen, P.; Mäkelä, A. Evaluation of Machine Learning Classifiers for Predicting Deep Convection. *J. Adv. Model. Earth Syst.* **2019**, *11*, 1784–1802. [CrossRef]
24. Zhou, K.; Zheng, Y.; Dong, W.; Wang, T. A Deep Learning Network for Cloud-to-Ground Lightning Nowcasting with Multisource Data. *J. Atmos. Ocean. Technol.* **2020**, *37*, 927–942. [CrossRef]
25. American Meteorological Society Overshooting Top—Glossary of Meteorology. Available online: https://glossary.ametsoc.org/wiki/Overshooting_top (accessed on 29 November 2023).
26. Bedka, K.M.; Allen, J.T.; Punge, H.J.; Kunz, M.; Simanovic, D. A Long-Term Overshooting Convective Cloud-Top Detection Database over Australia Derived from MTSAT Japanese Advanced Meteorological Imager Observations. *J. Appl. Meteorol. Climatol.* **2018**, *57*, 937–951. [CrossRef]
27. Khlopenkov, K.V.; Bedka, K.M.; Cooney, J.W.; Itterly, K. Recent Advances in Detection of Overshooting Cloud Tops From Longwave Infrared Satellite Imagery. *J. Geophys. Res. Atmos.* **2021**, *126*, e2020JD034359. [CrossRef]
28. Kim, M.; Im, J.; Park, H.; Park, S.; Lee, M.-I.; Ahn, M.-H. Detection of Tropical Overshooting Cloud Tops Using Himawari-8 Imagery. *Remote Sens.* **2017**, *9*, 685. [CrossRef]
29. Kim, M.; Lee, J.; Im, J. Deep Learning-Based Monitoring of Overshooting Cloud Tops from Geostationary Satellite Data. *GIScience Remote Sens.* **2018**, *55*, 763–792. [CrossRef]
30. Cavazos, T.; Turrent, C.; Lettenmaier, D.P. Extreme Precipitation Trends Associated with Tropical Cyclones in the Core of the North American Monsoon. *Geophys. Res. Lett.* **2008**, *35*, L21703. [CrossRef]

31. Mejia, J.F.; Douglas, M.W. Relationships between Moisture Surges and Mesoscale- to Large-Scale Convection from Multi-Year Satellite Imagery and North American Regional Reanalysis Data; AGU Fall Meeting Abstracts, A51F-0170. 2009. Available online: https://www.researchgate.net/publication/238027674_Relationships_between_Moisture_Surges_and_Mesoscale-_to_Large-Scale_Convection_from_Multi-year_Satellite_Imagery_and_North_American_Regional_Reanalysis_Data/citations#fullTextFileContent (accessed on 5 December 2023).
32. Ramos-Pérez, O.; Adams, D.K.; Ochoa-Moya, C.A.; Quintanar, A.I. A Climatology of Mesoscale Convective Systems in Northwest Mexico during the North American Monsoon. *Atmosphere* **2022**, *13*, 665. [CrossRef]
33. Forzieri, G.; Castelli, F.; Vivoni, E.R. Vegetation Dynamics within the North American Monsoon Region. *J. Clim.* **2011**, *24*, 1763–1783. [CrossRef]
34. Ramírez-López, A.K.; Carbajal, N.; Pineda-Martínez, L.F.; Tuxpan-Vargas, J. Cloud Climatology of Northwestern Mexico Based on MODIS Data. *Eur. J. Remote Sens.* **2023**, *56*, 2278066. [CrossRef]
35. INEGI. *Conjunto de Datos Vectoriales Escala 1:1 000 000. Unidades Climáticas*; INEGI: Aguascalientes, México, 2008.
36. Dávila Ortiz, R. Respuesta hidrológica en Zonas Urbanas Ante Eventos Severos de Tormenta Aplicando Modelación Numérica e Información SAR: Caso Los Mochis. Master's Thesis, Instituto Potosino de Investigación Científica y Tecnológica, A.C., San Luis Potosí, México, 2019.
37. INEGI. *Modelo Digital de Elevación de Alta Resolución Lidar*; INEGI: Aguascalientes, México, 2017.
38. León-Cruz, J.F.; Carbajal Henken, C.; Carbajal, N.; Fischer, J. Spatio-Temporal Distribution of Deep Convection Observed along the Trans-Mexican Volcanic Belt. *Remote Sens.* **2021**, *13*, 1215. [CrossRef]
39. Vargas, N.; Magaña, V. Climatic Risk in the Mexico City Metropolitan Area Due to Urbanization. *Urban Clim.* **2020**, *33*, 100644. [CrossRef]
40. Carbajal, N.; León-Cruz, J.F.; Pineda-Martínez, L.F.; Tuxpan-Vargas, J.; Gaviño-Rodríguez, J.H. Occurrence of Anticyclonic Tornadoes in a Topographically Complex Region of Mexico. *Adv. Meteorol.* **2019**, *2019*, e2763153. [CrossRef]
41. Romero Lankao, P. Water in Mexico City: What Will Climate Change Bring to Its History of Water-Related Hazards and Vulnerabilities? *Environ. Urban.* **2010**, *22*, 157–178. [CrossRef]
42. Eakin, H.; Lerner, A.M.; Manuel-Navarrete, D.; Hernández Aguilar, B.; Martínez-Canedo, A.; Tellman, B.; Charli-Joseph, L.; Fernández Álvarez, R.; Bojórquez-Tapia, L. Adapting to Risk and Perpetuating Poverty: Household's Strategies for Managing Flood Risk and Water Scarcity in Mexico City. *Environ. Sci. Policy* **2016**, *66*, 324–333. [CrossRef]
43. Schmit, T.J.; Griffith, P.; Gunshor, M.M.; Daniels, J.M.; Goodman, S.J.; Lebair, W.J. A Closer Look at the ABI on the GOES-R Series. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 681–698. [CrossRef]
44. Goodman, S.J.; Blakeslee, R.J.; Koshak, W.J.; Mach, D.; Bailey, J.; Buechler, D.; Carey, L.; Schultz, C.; Bateman, M.; McCaul, E.; et al. The GOES-R Geostationary Lightning Mapper (GLM). *Atmos. Res.* **2013**, *125–126*, 34–49. [CrossRef]
45. Platnick, S.; Meyer, K.G.; King, M.D.; Wind, G.; Amarasinghe, N.; Marchant, B.; Arnold, G.T.; Zhang, Z.; Hubanks, P.A.; Holz, R.E.; et al. The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 502–525. [CrossRef] [PubMed]
46. Rossow, W.B.; Schiffer, R.A. ISCCP Cloud Data Products. *Bull. Am. Meteorol. Soc.* **1991**, *72*, 2–20. [CrossRef]
47. Jergensen, G.E.; McGovern, A.; Lagerquist, R.; Smith, T. Classifying Convective Storms Using Machine Learning. *Weather Forecast.* **2020**, *35*, 537–559. [CrossRef]
48. Steinkruger, D.; Markowski, P.; Young, G. An Artificially Intelligent System for the Automated Issuance of Tornado Warnings in Simulated Convective Storms. *Weather Forecast.* **2020**, *35*, 1939–1965. [CrossRef]
49. Hosmer, D.; Lemeshow, S.; Sturdivant, R.X. Applied Logistic Regression. In *Wiley Series in Probability and Statistics*, 3rd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013; ISBN 978-1-118-54838-7.
50. Breiman, L. *Classification and Regression Trees*, 1st ed.; Routledge: New York, NY, USA, 1984; ISBN 978-1-315-13947-0.
51. Gagne, D.J.; McGovern, A.; Brotzge, J. Classification of Convective Areas Using Decision Trees. *J. Atmos. Ocean. Technol.* **2009**, *26*, 1341–1353. [CrossRef]
52. Vapnik, V. Pattern Recognition Using Generalized Portrait Method. *Autom. Remote Control* **1963**, *24*, 774–780.
53. Adrianto, I.; Trafalis, T.B.; Lakshmanan, V. Support Vector Machines for Spatiotemporal Tornado Prediction. *Int. J. Gen. Syst.* **2009**, *38*, 759–776. [CrossRef]
54. Trafalis, T.B.; Ince, H.; Richman, M.B. Tornado Detection with Support Vector Machines. In *Computational Science—ICCS 2003*; Sloot, P.M.A., Abramson, D., Bogdanov, A.V., Gorbachev, Y.E., Dongarra, J.J., Zomaya, A.Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 289–298.
55. Han, L.; Sun, J.; Zhang, W.; Xiu, Y.; Feng, H.; Lin, Y. A Machine Learning Nowcasting Method Based on Real-Time Reanalysis Data. *J. Geophys. Res. Atmos.* **2017**, *122*, 4038–4051. [CrossRef]
56. Sangiorgio, M.; Barindelli, S.; Guglieri, V.; Biondi, R.; Solazzo, E.; Realini, E.; Venuti, G.; Guariso, G. A Comparative Study on Machine Learning Techniques for Intense Convective Rainfall Events Forecasting. In *Theory and Applications of Time Series Analysis*; Valenzuela, O., Rojas, F., Herrera, L.J., Pomares, H., Rojas, I., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 305–317.
57. McCulloch, W.S.; Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]

58. Chen, H.; Chandrasekar, V.; Cifelli, R.; Xie, P. A Machine Learning System for Precipitation Estimation Using Satellite and Ground Radar Network Observations. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 982–994. [CrossRef]
59. Sobash, R.A.; Romine, G.S.; Schwartz, C.S. A Comparison of Neural-Network and Surrogate-Severe Probabilistic Convective Hazard Guidance Derived from a Convection-Allowing Model. *Weather Forecast.* **2020**, *35*, 1981–2000. [CrossRef]
60. Afzali Gorooh, V.; Kalia, S.; Nguyen, P.; Hsu, K.; Sorooshian, S.; Ganguly, S.; Nemani, R.R. Deep Neural Network Cloud-Type Classification (DeepCTC) Model and Its Application in Evaluating PERSIANN-CCS. *Remote Sens.* **2020**, *12*, 316. [CrossRef]
61. Lee, Y.; Han, D.; Ahn, M.-H.; Im, J.; Lee, S.J. Retrieval of Total Precipitable Water from Himawari-8 AHI Data: A Comparison of Random Forest, Extreme Gradient Boosting, and Deep Neural Network. *Remote Sens.* **2019**, *11*, 1741. [CrossRef]
62. Akbari Asanjan, A.; Yang, T.; Hsu, K.; Sorooshian, S.; Lin, J.; Peng, Q. Short-Term Precipitation Forecast Based on the PERSIANN System and LSTM Recurrent Neural Networks. *J. Geophys. Res. Atmos.* **2018**, *123*, 12543–12563. [CrossRef]
63. Leinonen, J.; Hamann, U.; Germann, U. Seamless Lightning Nowcasting with Recurrent-Convolutional Deep Learning. *Artif. Intell. Earth Syst.* **2022**, *1*, e220043. [CrossRef]
64. Hilburn, K.A.; Ebert-Uphoff, I.; Miller, S.D. Development and Interpretation of a Neural-Network-Based Synthetic Radar Reflectivity Estimator Using GOES-R Satellite Observations. *J. Appl. Meteorol. Climatol.* **2020**, *60*, 3–21. [CrossRef]
65. Lee, Y.; Kummerow, C.D.; Ebert-Uphoff, I. Applying Machine Learning Methods to Detect Convection Using Geostationary Operational Environmental Satellite-16 (GOES-16) Advanced Baseline Imager (ABI) Data. *Atmos. Meas. Tech.* **2021**, *14*, 2699–2716. [CrossRef]
66. Breiman, L. Bagging Predictors. *Mach Learn* **1996**, *24*, 123–140. [CrossRef]
67. Breiman, L. Pasting Small Votes for Classification in Large Databases and On-Line. *Mach. Learn.* **1999**, *36*, 85–103. [CrossRef]
68. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed.; O'Reilly: Sebastopol, CA, USA, 2019; ISBN 978-1-4920-3264-9.
69. Ramirez, S.; Lizarazo, I. Detecting and Tracking Mesoscale Precipitating Objects Using Machine Learning Algorithms. *Int. J. Remote Sens.* **2017**, *38*, 5045–5068. [CrossRef]
70. Ahijevych, D.; Pinto, J.O.; Williams, J.K.; Steiner, M. Probabilistic Forecasts of Mesoscale Convective System Initiation Using the Random Forest Data Mining Technique. *Weather Forecast.* **2016**, *31*, 581–599. [CrossRef]
71. Burke, A.; Snook, N.; Ii, D.J.G.; McCorkle, S.; McGovern, A. Calibration of Machine Learning–Based Probabilistic Hail Predictions for Operational Forecasting. *Weather Forecast.* **2020**, *35*, 149–168. [CrossRef]
72. Yao, H.; Li, X.; Pang, H.; Sheng, L.; Wang, W. Application of Random Forest Algorithm in Hail Forecasting over Shandong Peninsula. *Atmos. Res.* **2020**, *244*, 105093. [CrossRef]
73. Wolpert, D.H. Stacked Generalization. *Neural Netw.* **1992**, *5*, 241–259. [CrossRef]
74. NASA. *GOES-R Series Data Book*; GOES-R Series Program Office: Greenbelt, MD, USA, 2019.
75. La Fata, A.; Amato, F.; Bernardi, M.; D'Andrea, M.; Procopio, R.; Fiori, E. Horizontal Grid Spacing Comparison among Random Forest Algorithms to Nowcast Cloud-to-Ground Lightning Occurrence. *Stoch Env. Res Risk Assess* **2022**, *36*, 2195–2206. [CrossRef]
76. Rutledge, S.A.; Hilburn, K.A.; Clayton, A.; Fuchs, B.; Miller, S.D. Evaluating Geostationary Lightning Mapper Flash Rates Within Intense Convective Storms. *J. Geophys. Res. Atmos.* **2020**, *125*, e2020JD032827. [CrossRef]
77. Rudlosky, S.D.; Goodman, S.J.; Virts, K.S.; Bruning, E.C. Initial Geostationary Lightning Mapper Observations. *Geophys. Res. Lett.* **2019**, *46*, 1097–1104. [CrossRef]
78. Dávila Ortiz, R.; Tuxpan Vargas, J.; Velázquez Zapata, J.A. Identification of Deep Convection Clouds Using ABIGOES Data and Machine Learning Techniques: The Case of Los Mochis, Sinaloa, Mexico. In Proceedings of the 2023 IEEE Mexican International Conference on Computer Science (ENC), Guanajuato, Mexico, 11–13 September 2023. *in press*.
79. Kim, B.; Choi, S.Y.; Han, K.-Y. Integrated Real-Time Flood Forecasting and Inundation Analysis in Small–Medium Streams. *Water* **2019**, *11*, 919. [CrossRef]
80. da Silva Neto, C.P.; Alves Barbosa, H.; Assis Beneti, C.A. A Method for Convective Storm Detection Using Satellite Data. *Atmósfera* **2016**, *29*, 343–358. [CrossRef]
81. Maddox, R.A. Mesoscale Convective Complexes. *Bull. Am. Meteorol. Soc.* **1980**, *61*, 1374–1387. [CrossRef]
82. Anderson, C.J.; Arritt, R.W. Mesoscale Convective Complexes and Persistent Elongated Convective Systems over the United States during 1992 and 1993. *Mon. Wea. Rev.* **1998**, *126*, 578–599. [CrossRef]
83. Bedka, K.; Brunner, J.; Dworak, R.; Feltz, W.; Otkin, J.; Greenwald, T. Objective Satellite-Based Detection of Overshooting Tops Using Infrared Window Channel Brightness Temperature Gradients. *J. Appl. Meteorol. Climatol.* **2010**, *49*, 181–202. [CrossRef]
84. Machado, L.A.T.; Rossow, W.B. Structural Characteristics and Radiative Properties of Tropical Cloud Clusters. *Mon. Weather Rev.* **1993**, *121*, 3234–3260. [CrossRef]
85. Holle, R.L.; Murphy, M.J. Lightning in the North American Monsoon: An Exploratory Climatology. *Mon. Weather Rev.* **2015**, *143*, 1970–1977. [CrossRef]
86. Siqueira, J.R.; da Silva Marques, V. Tracking and Short-Term Forecasting of Mesoscale Convective Cloud Clusters over Southeast Brazil Using Satellite Infrared Imagery. *J. South. Hemisph. Earth Syst. Sci.* **2021**, *71*, 1–16. [CrossRef]
87. Liu, Z.; Min, M.; Li, J.; Sun, F.; Di, D.; Ai, Y.; Li, Z.; Qin, D.; Li, G.; Lin, Y.; et al. Local Severe Storm Tracking and Warning in Pre-Convection Stage from the New Generation Geostationary Weather Satellite Measurements. *Remote Sens.* **2019**, *11*, 383. [CrossRef]

88. Zhang, W.; Han, L.; Sun, J.; Guo, H.; Dai, J. Application of Multi-Channel 3D-Cube Successive Convolution Network for Convective Storm Nowcasting. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 1705–1710.
89. Bedka, K.M.; Khlopenkov, K. A Probabilistic Multispectral Pattern Recognition Method for Detection of Overshooting Cloud Tops Using Passive Satellite Imager Observations. *J. Appl. Meteorol. Climatol.* **2016**, *55*, 1983–2005. [CrossRef]
90. Kete, S.C.R.; Suprihatin; Tarigan, S.D.; Effendi, H. Land Use Classification Based on Object and Pixel Using Landsat 8 OLI in Kendari City, Southeast Sulawesi Province, Indonesia. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *284*, 012019. [CrossRef]

Graphical Abstract of the paper: Dávila-Ortiz, R., Carbajal-Pérez, J. N., Velázquez-Zapata, J. A., & Tuxpan-Vargas, J. (2024). Approximation of a Convective-Event-Monitoring System Using GOES-R Data and Ensemble ML Models. *Remote Sensing*, 16(4), 675.



2023 Mexican International Conference on Computer Science (ENC). Guanajuato, Guanajuato, Mexico

2023 Mexican International Conference on Computer Science (ENC) | 979-8-3503-9315-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ENC60556.2023.10508667

Identification of Deep Convection Clouds Using ABI-GOES Data and Machine Learning Techniques: The Case of Los Mochis, Sinaloa, Mexico

Rodrigo Dávila Ortiz
División de Geociencias Aplicadas
Instituto Potosino de Investigación Científica y Tecnológica
San Luis Potosí, México
ORCID: 0000-0003-4309-9269

José Tuxpan Vargas
División de Geociencias Aplicadas
Instituto Potosino de Investigación Científica y Tecnológica
San Luis Potosí, México
ORCID: 0000-0001-5167-0205

Juan Alberto Velázquez Zapata
Programa de Agua y Sociedad CONAHCYT-El Colegio de San Luis
San Luis Potosí, México
ORCID: 0000-0002-0208-9286

***Abstract*— This research addressed the implementation of a forecast framework, based on Machine Learning classification techniques, for the identification and monitoring of possible convective events associated with extreme precipitation. This methodology was applied in Los Mochis, Sinaloa, in Northeast Mexico, for this location presents intense convective activity and a high degree of vulnerability to floods and severe storms. The framework was constructed with a set of reference events from the Moderate-Resolution Imaging Spectroradiometer (MODIS) in which deep convective clouds were identified to generate the Target Labels set. Subsequently, a set of 12 predictor variables based on cloud optical properties is generated using data from the Advanced Baseline Imager (ABI) of the GOES-16 satellite. The Machine Learning models are Logistic Regression (LR) Random Forest (RF), which were selected for its simplicity, and their wide use in severe storms forecasting. The estimated Probability of Detection results were 0.8 in both RL and RF, while the False Alarm Ratio data was less than 0.2, indicating high modeling performance.**

Keywords*— *Deep Convection Clouds, Machine Learning, ABI-GOES, Logistic Regression, Random Forest.

I. INTRODUCTION

The monitoring and forecasting of meteorological conditions imply a significant challenge, especially in regions like Mexico where the instruments networks, such as meteorological radars or automatic stations, are insufficient [1]. In this regard, the use of satellite products represents an important and affordable alternative for monitoring convective hazards (i.e., weather events caused by the vertical movement of air masses, for example, lightning tornadoes, hail fall, severe storms, etc.).

On the other hand, Machine Learning (ML) and Deep Learning (DL) have been recognized as critical tools to solve problems in geosciences, including atmospheric sciences [2], [3]. Recently, advanced research has been conducted in the monitoring and forecasting of deep convective events using ML techniques, such as Logistic Regression (LR) [4], Random Forest (RF) and Decision Trees [5]–[9], and DL techniques [10]–[12].

In this study, LR and RF algorithms were used to identify potential deep convective clouds (vertically developed cloud systems driven by strong upward atmospheric motions) associated with the presence of severe storms, using data from the Advanced Baseline Imager (ABI) of the GOES R satellite constellation. Subsequently, the variables with the greatest weight were identified with a feature importance analysis. The results of this work set a precedent for the implementation of a convective hazards monitoring and forecasting framework with 5-min updates. The main novelty of this framework is that it can be scalable to any region of Mexico, since it only uses free access products and open-source software (Python 3).

II. MATERIALS AND METHODS

In this study, LR and RF models are used for the detection of deep convective clouds using data from the GOES-R Advanced Baseline Imager (ABI-GOES) as predictor variables, and data from the Moderate-Resolution Imaging Spectroradiometer (MODIS) for the generation of a set of target labels to classify two classes: cells with convective activity (CC) and cells with no convective activity (noCC). The databases, the study area, the classification models and the methods for the identification, validation, and evaluation of the models are described in this section.

A. Data

a) ABI-GOES Data

GOES-R ABI is a passive imaging radiometer with 16 spectral bands, including 2 visible channels, 4 near-infrared channels, and 10 infrared channels [13]. The ABI-GOES bands were used to generate 12 predictors, based on spectral properties of deep convective clouds. The temporal resolution for data collection is 5 min. The data are available in https://www.ncdc.noaa.gov/airs-web/search.

b) MODIS Data

MODIS is a passive imager mounted on both the Terra and Aqua sun-synchronous polar-orbiting satellites. This sensor provides pixel-level retrievals of cloud top properties and cloud optical properties [14]. For the generation of target labels, the parameters "Cloud_Top_Pressure" and "Cloud_Optical_Thickness" (related to the optical and microphysical properties of deep convective clouds) from the Collection 6 (C6) of MODIS products were collected. These data are available in https://www.ncdc.noaa.gov/airs-web/search.

B. Study area

Los Mochis is the main urban area of the municipality of Ahome (State of Sinaloa), Mexico. This city is located within the North American monsoon (NAM) core domain, so the region has a significant convective activity [15]–[17]. The meteorological characteristics and the socioeconomic conditions, produce a context of high vulnerability to severe storms and floods [18].

Figure 1 shows the location of Los Mochis between latitudes 25.6° - 26°, it is situated between the Gulf of California and the Sierra Madre Occidental; however, its topography is predominantly flat. This area is classified as Hot Desert climate (BWh), with a mean annual precipitation of 335 mm, which occurs mainly during a marked wet season between July and September. Its rainfall regime is strongly influenced by the presence of large-scale climatic phenomena such as El Niño–Southern Oscillation (ENSO) and hurricanes [18].

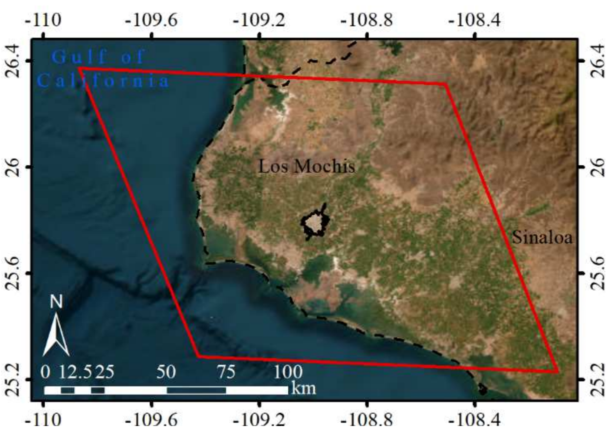


Fig. 1. Location of the study area. The red border corresponds to the special simulation domain, which has a size of 50 x 50 cells. The spatial range includes the municipal limit of Los Mochis, Sinaloa (black border) and an extended domain given the scale at which convection clouds are generated.

C. Models

LR is a statistical model used to predict a binary variable (here, CC or noCC) from one or more independent variables or predictors [19]. It works by fitting a logistic curve to the input data, which maps the input features to the output class probabilities (Fig. 2). The model parameters are estimated through maximum likelihood estimation, which finds the parameter values that best align with the observed data. Despite its simplicity, LR performs well for nonlinear problems in atmospheric science such as convection [20].

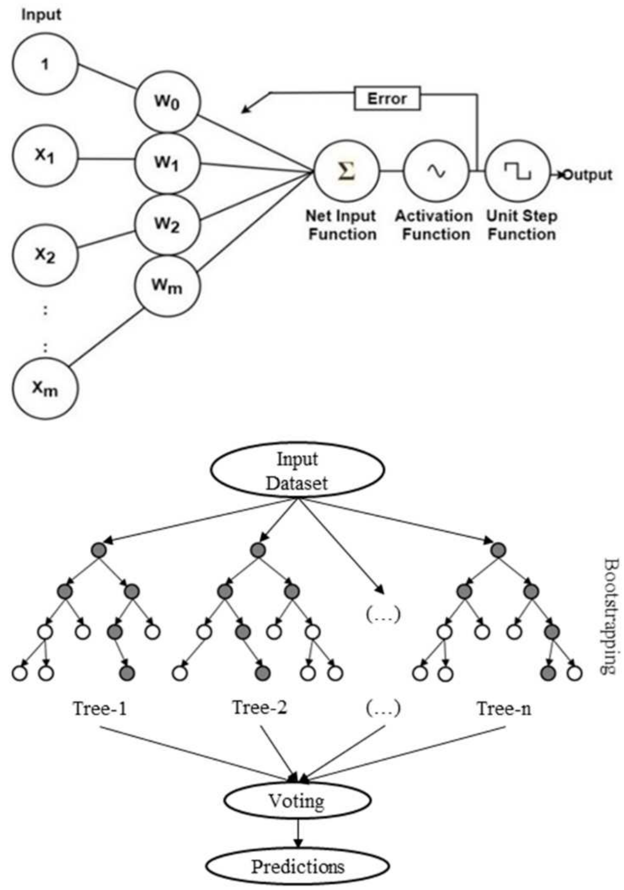


Fig. 2. Architecture of a Logistic Regression (upper) [24] and Random Forest (lower) [25].

RF is a nonparametric ensemble model based on the consensus of a collection of Decision Trees on different bootstrap samples, the outputs of the individual trees are then aggregated through voting to produce the final prediction (Fig. 2). The bootstrapping, that is, the training data resamples with replacement, allows the improvement of the model robustness and variance reduction through aggregation.

Along with LR, RF is one of the most widely used ML models in convective hazard forecasting studies [3], [4], [8], [21], [22], mostly due to its ability to capture non-linear association patterns between predictor and predictand, such as a convective storm system or precipitation [23]. Addressing the problem with both models allows the comparison of a low complexity model (LR), and a more robust model based on an assembly technique (RF)

D. Training and Testing datasets and Machine Learning Process

Data from the different ABI-GOES channels were used to generate 12 prediction variables, which in scientific literature are called "Interest Fields (Table I). In this study, the Interest Fields reported in the works of [26] and [6] were selected as predictors. On the other hand, the set of Target Labels was built using the MODIS C6 variables "Cloud_Optical_Thickness" (COT), parameter equivalent to Cloud Optical Depth) and

TABLE I . SUMMARY OF THE INPUT DATA USED FOR THE IMPLEMENTATION OF ML WORKFLOW AND POST-PROCESSING FILTER. THE INITIAL SELECTION OF PREDICTOR VARIABLES (INTEREST FIELDS) WAS TAKEN FROM [6].

ID	**Feature names**	**Type**	**Sensor**
CtT	Cloud-top temperature (**11.2 µm TB**)	Predictor	ABI
CtH01	Cloud-top height 01 (**6.2 - 11.2 µm**)	Predictor	ABI
CtH02	Cloud-top height 02 (**6.2 - 7.3 µm**)	Predictor	ABI
CtH03	Cloud-top height 03 (**13.3 - 11.2 µm**)	Predictor	ABI
CtG01	Cloud-top glaciation 01 (**12.3 - 11.2 µm**)	Predictor	ABI
CtG02	Cloud-top glaciation 02 (**8.6 - 11.2 µm**)	Predictor	ABI
CtG03	Cloud-top glaciation 03 (**8.6 - 11.2 µm**) - (**11.2 - 12.3 µm**)	Predictor	ABI
CtCrate	Cloud-top cooling rate (**11.2 µm time trend**)	Predictor	ABI
BTDrate	Brightness temperature difference rate (**6.2 - 11.2 µm time trend**)	Predictor	ABI
TChCtH02	Temporal changes in cloud-top height 02 (**6.2 - 7.3 µm time trend**)	Predictor	ABI
TChCtH03	Temporal changes in cloud-top height 03 (**13.3 - 11.2 µm time trend**)	Predictor	ABI
TChCtG03	Temporal changes in cloud-top glaciation ((**8.6 - 11.2 µm**) - (**11.2 - 12.3 µm**) **time trend**)	Predictor	ABI
CC_labels	Deep convective clouds labels	Target Variable	MODIS

"cloud_top_pression_1km" (CtP). The well-established classification of the International Satellite Cloud Climatology Project (ISCCP; Fig. 3) [27] was selected for the labeling of Convective Cells (CC). In particular, cells with values COT >23 and CtP < 440 mb are classified as deep convective clouds. The events selected for the construction of the training and testing datasets are listed in Table II. The size of the training and testing data sets are 15,000 and 20,000 cells respectively.

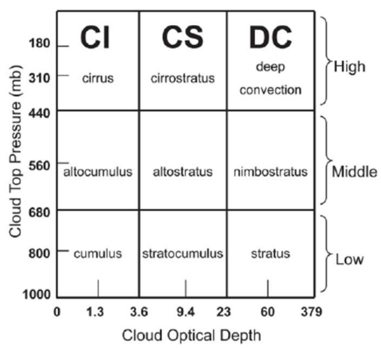


Fig. 3. ISCCP Cloud Classification in terms of Cloud Top Pressure and Cloud Optical Depth [27].

TABLE I. DATES AND TIMES OF CONVECTIVE EVENTS RECORDED IN ABI GOES-R IMAGERY USED TO CONSTRUCT THE TRAINING AND TESTING DATASETS.

Date	Time (UTC-00)	Usage
17-Jul-2018	18:30	Training data set
10-Aug-2018	20:50	
19-Sep-2018	18:30	
19-Sep-2018	20:00	
9-Aug-2019	21:15	
12-Aug-2021	18:10	
15-Aug-2018	18:00	Testing data set
22-Aug-2018	21:15	
29-Jul-2019	18:20	
8-Aug-2019	17:20	
22-Aug-2019	17:30	
23-Aug-2019	18:15	
14-Sep-2019	17:40	
30-Jun-2020	17:25	

A feature importance analysis was performed to estimate the degree of contribution of each of the interest fields on the outcome of a ML models. For LR, the relative importance of the predictor variables is given by the absolute value of the coefficients assigned to each Interest Field in the model training stage. In the case of RF, the Mean Decrease Impurity (MDI) is used as a measure of the degree of contribution of each predictor variable for the identification of the CC class. This method assesses the significance of each feature by quantifying its contribution to the impurity reduction, (here, Gini impurity), during the process of iterative feature-based splitting. In other words, the Gini impurity reduction is used to assess how each feature contributes to improve the purity of the divisions made in the tree construction process.

After selecting the predictors and scaling the data with the StandardScaler method (i.e., a function to scaling data from scikit-learn, a free software ML library of Python 3.0), a hyperparameter tuning was performed, the cross-validation strategy with 5 k-folds was used in the hyperparameter search process.

E. Metrics

To assess the performance of the different ML approaches, several classification metrics were calculated from the confusion matrix: Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI) and bias (BIAS).

$$\mathrm{POD} = TP \,/\, (TP + FN) \tag{1}$$

$$\mathrm{FAR} = FP \,/\, (FP + TP) \tag{2}$$

$$\mathrm{CSI} = TP \,/\, (TP + FP + FN) \tag{3}$$

$$\mathrm{BIAS} = (TP + FP) \,/\, (TP + FN) \tag{4}$$

where, TP is the number of true positives, FP indicates the number of false positives, FN is the number false negatives, and TN is the number of true negatives.

III. Results and Discussion

The relative feature importance analysis helps to reduce dimensionality, improve model efficiency, and mitigate the risk of overfitting. A lower dimensionality. dataset improves the model performance by removing noise from lower weight features, while computational cost is substantially reduced. In this sense, the training and testing datasets were delimited by reducing from twelve to six the number of features used as predictor variables. For this task, the criteria were the selection of values with the greatest relative importance in LR and RF (Table III). The Interest Fields with the highest degree of importance were CtT, CtH01, CtG01, CtH03, Ctg03 for LR, and CtH01, CtT, CtH03, CtH02, CtG01 for FR (Fig.4).

TABLE II. Interest Fields used as predictor variables after the Feature Importance Analysis.

Interest Fields used as predictor variable
Cloud-top temperature (CtT)
Cloud-top height 01 (CtH01)
Cloud-top height 02 (CtH02)
Cloud-top height 03 (CtH03)
Cloud-top glaciation 01 (CtG01)
Cloud-top glaciation 03 (CtG03)

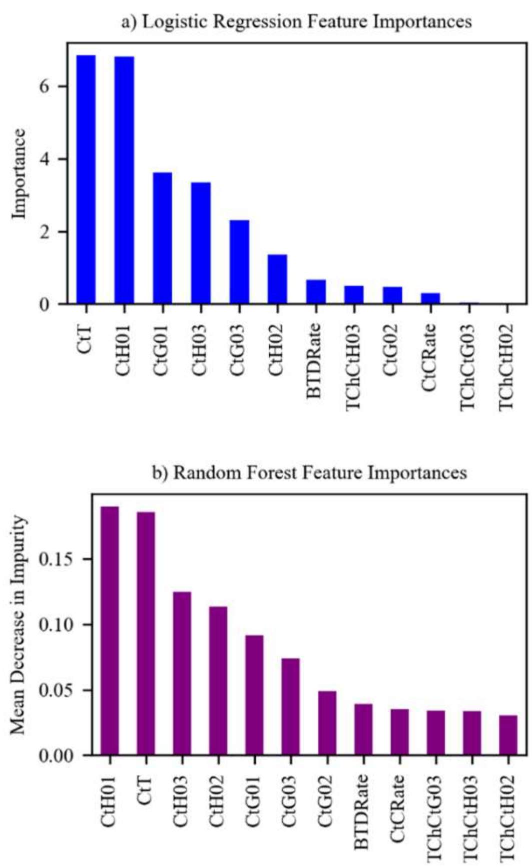


Fig. 4. Analysis of the feature importance using the absolute value of the coefficients obtained in the training stage of the LR model (a) and the MDI estimated for the 12 Interest Fields with the RF model (b). A high value of the weight of the coefficients and of the MDI value is interpreted as an indicator of the degree of contribution of a variable.

In the first case, CtT, as its name indicates, directly corresponds to the brightness temperature measured at the top of the cloud. This agrees with the typical properties of a deep convective cloud, where at higher altitudes, the temperature at the cloud top is colder than the surrounding environment. For this reason, these kind of clouds reach significant heights in the atmosphere, where temperatures are much lower due to the decreasing atmospheric pressure with altitude. The cold cloud top is the result of the adiabatic cooling process, where rising air expands and cools as it moves upward through the atmosphere. These findings agree with the results reported by references [4], [6], [26] and [9], which also identified the CtT among the most contributing Interest Fields.

Figure 5 shows the distribution of the CtT values for each of the classes for the entire reference data set (the training data set has 9,736 CC and 5,264 noCC cells, while the testing data set has 14,380 and 5,620 CC and noCC cells respectively). From this figure, it can be seen that the sets of CC and noCC cells are statistically differentiable for this variable (which is a desirable condition for classification tasks), showing a clear relationship between the coldest CtT values and the presence of CC cells. In this sense, reference [28] lists some cloud-top temperature thresholds reported in studies related to the identification of deep convection clouds, for example, 241 and 221 K [29], 221 K [30], 255 and 206 K [31], and 225 K [32], even reference [28] uses a threshold of 233 K as criteria for their convective storm detection algorithm.

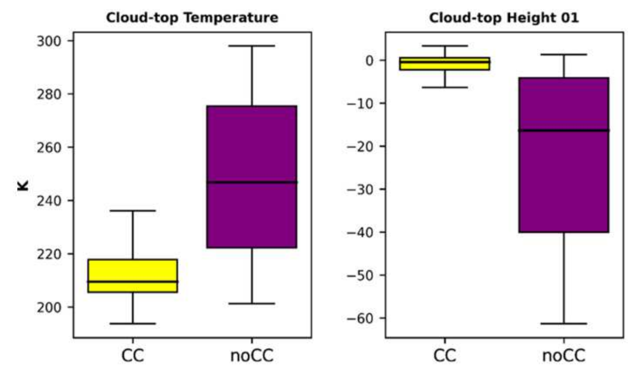


Fig. 5. Box plots of the most important input variables identified after the importance feature analysis, Cloud-top Temperature (CtT) and Cloud-top height 01 (CtH01). In this figure, the distribution of this variables of all reference data was used for CC detection models.

The Second Interest Field in order of importance is CtH01, which represents the difference between channels 6.2 and 11.2 µm. This value is used to determine lower stratospheric moisture, and it is positive when water vapor is present above the cloud tops [9], which is an indicator of the presence of an Overshooting top, that is, a dome-like protrusion shooting out of the top of the anvil of a thunderstorm and into the lower stratosphere [33]. These clouds are typically associated with thunderstorms and intense convection. In this case, its distribution is concentrated in values close to 0 for the case of CC cells, while the noCC values range between 0 and -60. The distributions of the CC and noCC sets are statistically differentiable in the same way as CtT.

Figures 6 and 7 show the results of the hyperparameter tuning for LR and RF respectively. In the case of LR, the maximum value of the accuracy score was 0.82, reached with a configuration C = 0.01 and the liblinear solver. The parameter C is the inverse of the regularization strength, where smaller values of C correspond to stronger regularization. Regularization prevents overfitting by penalizing large coefficient values, thus favoring simpler models. When C is large, the model aims to minimize misclassification of training data, potentially leading to complex models that fit noise. On the other hand, smaller C values emphasize coefficient regularization, promoting models that might better generalize to new data. The liblinear solver is a numerical optimization algorithm employed to find the optimal coefficients that minimize the logistic loss function.

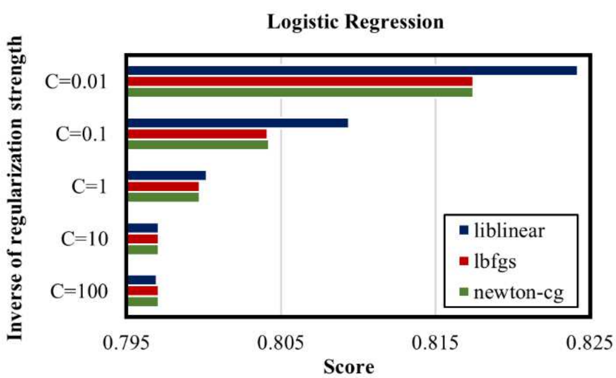


Fig. 6. Results of hyperparameter tuning for LR. The best fit parameter set found is: Inverse of regularization strength C = 0.01 and solver: liblinear.

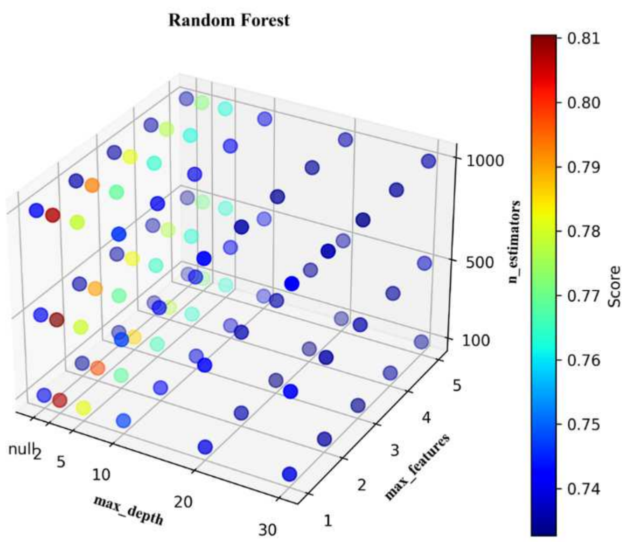


Fig. 7. Results of hyperparameter tuning for RF. The best fit parameter set is: number of trees (n_estimators) = 500, maximum depth of the tree (max_depth) = 2, and random split predictor variables (max_features) = 1.

On the other hand, RF reaches a similar value of 0.81 with a configuration, n_estimators = 500, max_depth = 2, and max_features = 1. In RF context, n_estimators determine the number of decision trees (DT) included in the ensemble, the parameter max_depth, specifies the maximum depth that each individual DT can grow to during training stage, and max_features controls the number of features that are considered for splitting at each node while constructing individual DT.

Regarding the performance metrics, an approximate POD value of 0.8 is reached in both LR and RF, and FAR values lower than 0.2 were obtained (Fig. 8). This result evidences a good balance between the probability of detection of the generation of a deep convective event, with a good level of certainty that they are not false alarms. In this sense, reported values in other works can be mentioned as a reference; for example, values of POD $\approx$ 0.8 and FAR $\approx$ 0.2 were reported [6], other works obtained values of POD $\approx$ 0.82 and FAR $\approx$ 0.37[26], and POD $\approx$ 0.87 and FAR $\approx$ 0.2 - 0.36 [4]. Other useful metric in binary forecasting is CSI, which provides a measure of the proportion of correct forecasts (both hits and correct rejections) to all the possible forecasts, including both correct forecasts and errors. It ranges from 0 to 1, with a value of 1 representing a perfect forecast and 0 indicating no skill. Our results show CSI values in LR and RF close to 0.7, which agree the results obtained by [34] CSI $\approx$ 0.74 - 0.8 and [35] CSI $\approx$ 0.65.

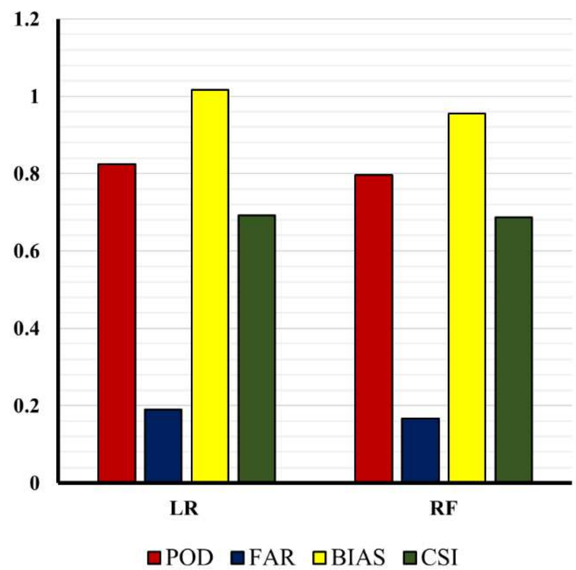


Fig. 8. Assessment of the testing metrics POD, FAR, CSI and BIAS for the LR and RF models.

According to reference [36] the BIAS metric ranges from 0 to infinity, and allows evaluating whether the forecasting method tends to underestimate the CC occurrence (BIAS<1) or to overestimate it (BIAS>1). BIAS=0 corresponds to a perfect forecast. Our results show that the BIAS value is higher than 1 with LR, while with RF the value is lower; however, both models obtained a very close value to perfect forecast.

Finally, Fig. 9 shows the simulation of the event 2019-234 - 17:30 (August 22, 2019) with LR y RF. When comparing the event simulation with the temperature map at the top of the cloud, it can be seen a good correspondence between the spatial distribution of the CC cells and the coldest areas of the imagery. For its part, LR tends to identify a larger portion of CC cells at

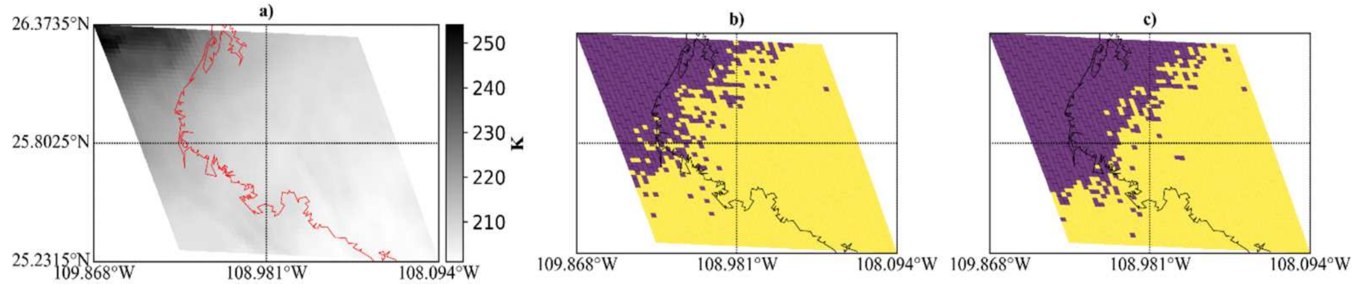


Fig. 9. Cloud-top brightness temperature map of the Los Mochis 2019-234 - 17:30 event (August 22, 2019) obtained from the ABI Band 14 (a), and simulation of the event with LR (b) and RF (c) models. The yellow regions indicate CC simulate pixels and the purple correspond with the noCC pixels.

the edge of the cloud. The transition zones between both classes are difficult to simulate due to the diffusion effect of the optical properties of the cloud that the models use to determine the CC occurrence.

Future work will include an expanded selection of ML models for the detection of potential deep convective clouds, as well as the use of model assembly strategies such as Voting and Stacking. Also. DL techniques such as Recurrent Neural Networks (RNN) and Convolutional Networks (CNN) will be tested. In addition, the incorporation of additional study areas will be useful to determine the effect of each of the predictor variables in regions with different convection mechanisms and atmospheric dynamics.

IV. CONCLUSIONS

A machine learning framework was implemented based on LR and RF models to identify probable severe convective events. The design was implemented in Los Mochis (Sinaloa, Mexico) due to its high vulnerability to extreme precipitation events.

The simulation experiments identified, in both models, six predictor variables with a higher contribution degree, a result that is consistent with other recent works for the identification of fields of interest estimated from ABI GOES sensor observations.

According to the metrics obtained, the identification of convective events from satellite data, ML-based models and model optimization represents a reliable strategy that open the possibility of generating autonomous and intelligent systems for monitoring and tracking convective events on a larger spatial scale, taking advantage of the great availability of temporal information at near-real-time levels.

REFERENCES

[1] J. F. León-Cruz, C. Carbajal Henken, N. Carbajal, and J. Fischer, “Spatio-Temporal Distribution of Deep Convection Observed along the Trans-Mexican Volcanic Belt,” Remote Sensing, vol. 13, no. 6, Art. no. 6, Jan. 2021, doi: 10.3390/rs13061215.

[2] M. Reichstein et al., “Deep learning and process understanding for data-driven Earth system science,” Nature, vol. 566, no. 7743, Art. no. 7743, Feb. 2019, doi: 10.1038/s41586-019-0912-1.

[3] A. McGovern et al., “A Review of Machine Learning for Convective Weather,” Artificial Intelligence for the Earth Systems, vol. 1, no. aop, pp. 1–61, May 2023, doi: 10.1175/AIES-D-22-0077.1.

[4] J. R. Mecikalski, J. K. Williams, C. P. Jewett, D. Ahijevych, A. LeRoy, and J. R. Walker, “Probabilistic 0–1-h Convective Initiation Nowcasts that Combine Geostationary Satellite Observations and Numerical Weather Prediction Model Data,” Journal of Applied Meteorology and Climatology, vol. 54, no. 5, pp. 1039–1059, May 2015, doi: 10.1175/JAMC-D-14-0129.1.

[5] D. J. Gagne, A. McGovern, and J. Brotzge, “Classification of Convective Areas Using Decision Trees,” Journal of Atmospheric and Oceanic Technology, vol. 26, no. 7, pp. 1341–1353, Jul. 2009, doi: 10.1175/2008JTECHA1205.1.

[6] S. Lee, H. Han, J. Im, E. Jang, and M.-I. Lee, “Detection of deterministic and probabilistic convection initiation using Himawari-8 Advanced Himawari Imager data,” Atmospheric Measurement Techniques, vol. 10, no. 5, pp. 1859–1874, May 2017, doi: 10.5194/amt-10-1859-2017.

[7] H. Han et al., “Detection of Convective Initiation Using Meteorological Imager Onboard Communication, Ocean, and Meteorological Satellite Based on Machine Learning Approaches,” Remote Sensing, vol. 7, no. 7, Art. no. 7, Jul. 2015, doi: 10.3390/rs70709184.

[8] G. E. Jergensen, A. McGovern, R. Lagerquist, and T. Smith, “Classifying Convective Storms Using Machine Learning,” Weather and Forecasting, vol. 35, no. 2, pp. 537–559, Apr. 2020, doi: 10.1175/WAF-D-19-0170.1.

[9] M. Kim, J. Im, H. Park, S. Park, M.-I. Lee, and M.-H. Ahn, “Detection of Tropical Overshooting Cloud Tops Using Himawari-8 Imagery,” Remote Sensing, vol. 9, no. 7, Art. no. 7, Jul. 2017, doi: 10.3390/rs9070685.

[10] V. Afzali Gorooh et al., “Deep Neural Network Cloud-Type Classification (DeepCTC) Model and Its Application in Evaluating PERSIANN-CCS,” Remote Sensing, vol. 12, no. 2, Art. no. 2, Jan. 2020, doi: 10.3390/rs12020316.

[11] Y. Lee, D. Han, M.-H. Ahn, J. Im, and S. J. Lee, “Retrieval of Total Precipitable Water from Himawari-8 AHI Data: A Comparison of Random Forest, Extreme Gradient Boosting, and Deep Neural Network,” Remote Sensing, vol. 11, no. 15, Art. no. 15, Jan. 2019, doi: 10.3390/rs11151741.

[12] Q. Liu et al., “Daytime Rainy Cloud Detection and Convective Precipitation Delineation Based on a Deep Neural Network Method Using GOES-16 ABI Images,” Remote Sensing, vol. 11, no. 21, Art. no. 21, Jan. 2019, doi: 10.3390/rs11212555.

[13] T. J. Schmit, P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebair, “A Closer Look at the ABI on the GOES-R Series,” Bulletin of the American Meteorological Society, vol. 98, no. 4, pp. 681–698, Apr. 2017, doi: 10.1175/BAMS-D-15-00230.1.

[14] S. Platnick et al., “The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua,” IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 1, pp. 502–525, Jan. 2017, doi: 10.1109/TGRS.2016.2610522.

[15] T. Cavazos, C. Turrent, and D. P. Lettenmaier, “Extreme precipitation trends associated with tropical cyclones in the core of the North American monsoon,” Geophysical Research Letters, vol. 35, no. 21, 2008, doi: 10.1029/2008GL035832.

[16] J. F. Mejia and M. W. Douglas, “Relationships between Moisture Surges and Mesoscale- to Large-Scale Convection from Multi-year Satellite

Imagery and North American Regional Reanalysis Data," vol. 2009, pp. A51F-0170, Dec. 2009.

[17] O. Ramos-Pérez, D. K. Adams, C. A. Ochoa-Moya, and A. I. Quintanar, "A Climatology of Mesoscale Convective Systems in Northwest Mexico during the North American Monsoon," Atmosphere, vol. 13, no. 5, Art. no. 5, May 2022, doi: 10.3390/atmos13050665.

[18] R. Dávila Ortiz, "Respuesta hidrológica en zonas urbanas ante eventos severos de tormenta aplicando modelación numérica e información SAR: Caso Los Mochis," masterThesis, 2019. Accessed: Apr. 08, 2023. [Online]. Available: https://repositorio.ipicyt.edu.mx///handle/11627/5259

[19] D. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, Third. in Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 2013. doi: 10.1002/9781118548387.

[20] P. Ukkonen and A. Mäkelä, "Evaluation of Machine Learning Classifiers for Predicting Deep Convection," Journal of Advances in Modeling Earth Systems, vol. 11, no. 6, pp. 1784–1802, 2019, doi: 10.1029/2018MS001561.

[21] A. La Fata, F. Amato, M. Bernardi, M. D'Andrea, R. Procopio, and E. Fiori, "Cloud-to-Ground lightning nowcasting using Machine Learning," in 2021 35th International Conference on Lightning Protection (ICLP) and XVI International Symposium on Lightning Protection (SIPDA), Sep. 2021, pp. 1–6. doi: 10.1109/ICLPandSIPDA54065.2021.9627428.

[22] D. Steinkruger, P. Markowski, and G. Young, "An Artificially Intelligent System for the Automated Issuance of Tornado Warnings in Simulated Convective Storms," Weather and Forecasting, vol. 35, no. 5, pp. 1939–1965, Oct. 2020, doi: 10.1175/WAF-D-19-0249.1.

[23] S. Ramirez and I. Lizarazo, "Detecting and tracking mesoscale precipitating objects using machine learning algorithms," International Journal of Remote Sensing, vol. 38, no. 18, pp. 5045–5068, Sep. 2017, doi: 10.1080/01431161.2017.1323280.

[24] R. Torres, O. Ohashi, and G. Pessin, "A Machine-Learning Approach to Distinguish Passengers and Drivers Reading While Driving," *Sensors (Basel)*, vol. 19, no. 14, p. 3174, Jul. 2019, doi: 10.3390/s19143174.

[25] H. Sahour, V. Gholami, J. Torkaman, M. Vazifedan, and S. Saeedi, "Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings," Environ Earth Sci, vol. 80, no. 22, p. 747, Nov. 2021, doi: 10.1007/s12665-021-10054-5

[26] D. Han, J. Lee, J. Im, S. Sim, S. Lee, and H. Han, "A Novel Framework of Detecting Convective Initiation Combining Automated Sampling, Machine Learning, and Repeated Model Tuning from Geostationary Satellite Data," Remote Sensing, vol. 11, no. 12, Art. no. 12, Jan. 2019, doi: 10.3390/rs11121454.

[27] W. B. Rossow and R. A. Schiffer, "ISCCP Cloud Data Products," Bulletin of the American Meteorological Society, vol. 72, no. 1, pp. 2–20, Jan. 1991, doi: 10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2.

[28] C. P. da Silva Neto, H. Alves Barbosa, and C. A. Assis Beneti, "A method for convective storm detection using satellite data," Atmósfera, vol. 29, no. 4, pp. 343–358, Oct. 2016, doi: 10.20937/ATM.2016.29.04.05.

[29] R. A. Maddox, "Mesoscale Convective Complexes," Bulletin of the American Meteorological Society, vol. 61, no. 11, pp. 1374–1387, 1980.

[30] C. J. Anderson and R. W. Arritt, "Mesoscale Convective Complexes and Persistent Elongated Convective Systems over the United States during 1992 and 1993," Mon. Wea. Rev., vol. 126, no. 3, pp. 578–599, Mar. 1998, doi: 10.1175/1520-0493(1998)126<0578:MCCAPE>2.0.CO;2.

[31] L. a. T. Machado and W. B. Rossow, "Structural Characteristics and Radiative Properties of Tropical Cloud Clusters," Monthly Weather Review, vol. 121, no. 12, pp. 3234–3260, Dec. 1993, doi: 10.1175/1520-0493(1993)121<3234:SCARPO>2.0.CO;2.

[32] K. Bedka, J. Brunner, R. Dworak, W. Feltz, J. Otkin, and T. Greenwald, "Objective Satellite-Based Detection of Overshooting Tops Using Infrared Window Channel Brightness Temperature Gradients," Journal of Applied Meteorology and Climatology, vol. 49, no. 2, pp. 181–202, Feb. 2010, doi: 10.1175/2009JAMC2286.1.

[33] W. E. Shenk, "Cloud Top Height Variability of Strong Convective Cells," Journal of Applied Meteorology and Climatology, vol. 13, no. 8, pp. 917–922, Dec. 1974, doi: 10.1175/1520-0450(1974)013<0917:CTHVOS>2.0.CO;2.

[34] C. Muñoz, L.-P. Wang, and P. Willems, "Enhanced object-based tracking algorithm for convective rain storms and cells," Atmospheric Research, vol. 201, pp. 144–158, Mar. 2018, doi: 10.1016/j.atmosres.2017.10.027.

[35] A. Su, H. Li, L. Cui, and Y. Chen, "A Convection Nowcasting Method Based on Machine Learning," Advances in Meteorology, vol. 2020, p. e5124274, Jan. 2020, doi: 10.1155/2020/5124274.

[36] J. R. Siqueira, V. da S. Marques, J. R. Siqueira, and V. da S. Marques, "Tracking and short-term forecasting of mesoscale convective cloud clusters over southeast Brazil using satellite infrared imagery," JSHESS, vol. 71, no. 1, pp. 1–16, Feb. 2021, doi: 10.1071/ES19050.

EL COLEGIO DE SAN LUIS, A.C.

San Luis Potosí, San Luis Potosí, a 21 de marzo de 2024

Dr. Rodrigo Dávila Ortiz
Presente.-

Por este medio le informo que el libro *Riesgos y desastres relacionados con agua: transformación del territorio, inundaciones y contaminación,* coordinado por Edgar Talledos Sánchez y Juan Alberto Velázquez Zapata, donde aparece el capítulo "La producción del riesgo ante eventos de inundación en Los Mochis, Sinaloa", de su autoría, la de José Tuxpan Vargas y del Dr. Velázquez Zapata, fue dictaminado favorablemente por el método de doble ciego y aprobado para publicarse en el fondo editorial de El Colegio de San Luis.

Sin más que agregar, le saludo cordialmente.

Ing. Jorge Herrera Patiño
Jefe de la Unidad de Publicaciones
El Colegio de San Luis
jorge.herrera@colsan.edu.mx

C.c.p. Archivo UP

La producción del riesgo en ante eventos de inundación en Los Mochis, Sinaloa

Resumen

La ciudad de Los Mochis se localiza en la cuenca del río Fuerte y dentro del distrito de Riego 075 en Sinaloa, México. La ciudad ha sido afectada, desde su fundación, por eventos de inundación, algunos de ellos muy severos como los acontecidos en 2008 y en 2018, en los que casi la totalidad de la ciudad ha sido cubierta por agua, provocando la pérdida del patrimonio de la población, daños a la infraestructura y la evacuación de miles de residentes. Si bien la ubicación geográfica de la ciudad, cercana a la costa del Golfo de California, la hace susceptible ser afectada por ciclones tropicales y otros fenómenos meteorológicos, este estudio analiza el contexto histórico de configuración de la ciudad como el principal factor de producción de riesgo la ciudad. Además, el estudio muestra, con base en el análisis del desastre de septiembre de 2018, el contexto actual de vulnerabilidad de la población.

Palabras clave: *Inundaciones, Los Mochis, Sinaloa*

Abstract

The city of Los Mochis is located in the Río Fuerte basin and within the Irrigation District in Sinaloa, Mexico. The city has been affected, since its foundation, by flood events, some of them very severe such as those that occurred in 2008 and 2018, in which almost the entire city has been covered by water, causing the loss of the heritage of the population, damage to infrastructure and the evacuation of thousands of residents. Although the geographic location of the city, close to the coast of the Gulf of California, makes it susceptible to being affected by tropical cyclones and other meteorological phenomena, this study analyzes the historical context of the city's configuration as the main factor of risk production the city. In addition, the study shows, based on the analysis of the September 2018 disaster, the current context of vulnerability of the population.

Keywords: *Floods, Los Mochis, Sinaloa*

1.Introducción

De acuerdo con la Organización Meteorológica Mundial, se han registrado 8,835 desastres a nivel global entre los años 1970 a 2012, de los cuales las inundaciones constituyen el 44% de estos eventos (Golnaraghi, Etienne, Guha-Sapir, & Below, 2014), siendo así el desastre más frecuente a nivel mundial.

La explicación a la causa de los desastres por inundaciones suele darse evaluando la cantidad de precipitación en el lugar afectado o en las condiciones geográficas, mientras que las consecuencias se estiman con el recuento de daños a la infraestructura y el número de afectados, obviando muchas veces las condiciones de riesgo y vulnerabilidad que dieron origen a dichas afectaciones. Ejemplo de este enfoque basado en el aspecto de geografía física es el estudio de Arreguín-Cortés *et al.* (2016) en el cual estiman que 162 000 km^2 del territorio mexicano son susceptibles de inundarse identificando como principal causa un problema de ordenamiento territorial.

Autores como Blaikie, Cannon, Davis, & Wisner (1996) afirman que cuando se presenta un desastre está implicado un fenómeno físico, o amenaza de origen geológica o hidrometeorológico (entre otras), sin embargo, son las condiciones políticas, sociales y económicas del entorno la principal causa del desastre. En este contexto, estos autores identifican la vulnerabilidad como “las características de una persona o grupo desde el punto de vista de su capacidad para anticipar, sobrevivir, resistir y recuperarse del impacto de una amenaza natural”. Por lo tanto, una amenaza puede afectar a la población en menor o mayor grado, según las condiciones preexistentes de la sociedad y los recursos (económicos, institucionales, etc.) de los que ésta disponga para recuperarse.

El desastre generalmente se ve como una disrupción de la normalidad, esto es un evento que se presenta súbitamente y que afecta el funcionamiento ordenado de una comunidad. Un enfoque para analizar el desastre, es estudiarlo como un proceso que se va construyendo con la influencia de diversos factores sociales. Entre estos se encuentran factores socioeconómicos como la marginalidad, la densidad de población, la pobreza, la falta de sistemas de alerta ante amenazas y los contextos de vulnerabilidad de la población (Rodríguez Esteves, 2007). El estudio del proceso de conformación del desastre tiene también una componente histórica, como el propuesto por Ley García & Calderón Aragón (2008), quienes estudian la producción del riesgo y la vulnerabilidad en la ciudad de Mexicali (Baja California) en las tres décadas a partir de su fundación. En este estudio, se identifica las condiciones de alejamiento de la ciudad respecto al resto del país como un factor que propició la ausencia de políticas nacionales en la región, lo que llevó a que los extranjeros se apropiaran de los medios de producción y comercio. Este contexto llevó a que Mexicali se convirtiera en un espacio frágil ante los peligros naturales y con población poco resiliente.

En base a los conceptos anteriores, este estudio evalúa la producción del riesgo ante inundaciones en la ciudad de Los Mochis, considerando los componentes físicos y sociales. El estudio presenta el contexto histórico de la fundación de la ciudad y las condiciones que las llevaron a convertirse en un importante polo de desarrollo agrícola. Además, se evalúa en contexto físico de la ciudad, entre ellos la red hidrológica natural y la construcción de infraestructura hidráulica que conformaron las condiciones actuales de Los Mochis. A continuación, se presenta un caso reciente e importante de inundación acontecido en 2018 y se evalúan las condiciones que llevaron a dicho desastre. Finalmente, se presentan las conclusiones que cierran el trabajo.

2. Contexto histórico de la ciudad de los Mochis

En la actualidad, la ciudad de Los Mochis cabecera municipal de Ahome, en el Estado de Sinaloa, es considerada como un referente de modernidad y crecimiento, no solo a nivel estatal sino en todo el país. Con frecuencia, esta ciudad se considera como la tercera más importante del estado debido a su gran producción agrícola (CONAGUA & SEMARNAT, 2017), posicionándola como uno de los emporios agrícolas más grandes del país.

Algunos cronistas e historiadores consideran el origen de este centro urbano previo a los tiempos de la conquista; por ejemplo José Armando Infante identifica tres momentos fundamentales en la formación de esta ciudad a los cuales denomina como “las tres fundaciones de Los Mochis” (Infante, 2011), no obstante, acotaremos el inicio de esta historia en septiembre de 1872, fecha en que la gran mayoría considera comienza la fundación de los Mochis con la llegada del ingeniero civil estadounidense Albert K. Owen quien llegó al valle del Río Fuerte y a la bahía de Ohuira como parte de un equipo de exploración de una empresa americana de ferroviaria en busca de un puerto en la costa del noroeste que fungiera como terminal de línea ferroviaria que correría hasta Nueva York con el fin de agilizar el comercio entre Estados Unidos y los países asiáticos (Instituto Latinoamericano de la Comunicación Educativa, 2021).

Se dice que en cuanto el ingeniero Owen arribó a la bahía de Ohuira quedó maravillado y convencido de que ese lugar sería donde se construiría “la ciudad del futuro” y fue así que comenzó con un desarrollo portuario y urbanístico, al estilo de San Francisco o Nueva York, diseñada con amplias calles y avenidas (Infante, 2011). Fue así que años después Owen junto con un grupo de socialistas utópicos crearon la ciudad de *Pacific City* (la cual es hoy en día conocido como Topolobampo) bajo los principios de socialismo utópico, además comenzaron con la construcción de los primeros canales y acueductos para derivar agua proveniente del Río Fuerte y aprovechar las fértiles tierras que se extendían por todo Valle del Fuerte (mochisonline, 2019). Este hecho, aunque poco conocido fue importante, según se describe en la reseña de Quintana Navarrete (2015), “este diseño socioeconómico guío a los pioneros estadounidenses que, entusiasmados por las ideas utópicas de Owen y por las perspectivas de una vida feliz y próspera en tierras mexicanas, desembarcaron en las playas de Sinaloa en noviembre de 1886”. Entre la gran cantidad de personajes que arribaron a costas mexicanas atraídos por el proyecto de Owen se encontraba Benjamin Francis Johnston quien fue una figura de vital importancia durante los siguientes años.

Pasaron 30 años y la utopía de Owen comenzaba a desmoronarse debido a diversas condiciones adversas, por ejemplo, las costas de arenosas y salinas de Topolobampo no eran aptas para la agricultura, además de enfrentar rencillas políticas y un brote de paludismo (Quintana Navarrete, 2015). No obstante, algunos pobladores se establecieron en Topolobampo continuando con la construcción de canales de riego, una parte de los pobladores regresó a Estados Unidos, pero el mayor grueso de colonos se estableció en dos poblados que se encontraban al sur sobre la trayectoria de una de las principales obras hidráulicas que a la actualidad sigue sirviendo para derivar importantes volúmenes de agua para el riego de la gran cantidad de campos de cultivo de esa región, el “Canal Taxtes”, del cual se hablará a detalle en las secciones posteriores (H. Ayuntamiento de Ahome, 2018). Estos dos poblados “*The Public Farm*” o el “Público” como mejor se le conocía en ese tiempo y “*El Plat*”, se situaron en lo que hoy en día se conoce como Los Mochis, aunque no fue sino hasta el 1903, que el gobernador Francisco Cañedo decreta el cambio de nombre (Infante, 2011).

Durante esos años, a la par de que se continuaba con el proyecto ferroviario, Benjamin Francis Johnston comienza a explotar recursos como la caña de azúcar, crenado en sociedad Edward Lycan y Zacarías Ochoa "el Águila Sugar & Refining Company", que después cambia de nombre a "Sinaloa Sugar Company" y luego a "United Sugar Company" la cual sigue operando a la fecha con el nombre de "Compañía Azucarera de Los Mochis". Este ingenio azucarero hasta la fecha es considerado como uno de los más importantes a nivel nacional, y en su tiempo fue una pieza elemental para la consolidación del gran desarrollo que vería esa joven ciudad durante los años venideros y un importante polo de crecimiento demográfico (mochisonline, 2019) que para 1910 contaba con 1,188 habitantes (H. Ayuntamiento de Ahome, 2018).

Con el auge de industria cañera el sistema de irrigación y las concesiones federales pasaron a manos del Johnston, quien amplió y modernizó los campos de cultivo y la red de riego con agua proveniente del Río Fuerte (Instituto Latinoamericano de la Comunicación Educativa, 2021).

En 1916 se creó el municipio de Ahome y desde 1935 la cabecera municipal de este último se ubica en la ciudad de Los Mochis (mochisonline, 2019).

Otro hecho de gran relevancia fue el decreto de instauración del Distrito de Riego 075, "Río Fuerte" en cual se estableció sobre centro agrícola perteneciente al Valle del Fuerte al cual pertenece la ciudad de Los Mochis. Éste se creó mediante acuerdo presidencial publicado en el diario oficial de la federación el 21 de agosto de 1951. La fecha de inicio de operación del distrito fue el año de 1956 con la terminación de la primera etapa de construcción de la presa Miguel Hidalgo y fue transferida a los usuarios en 1992 (Barrera Pérez, 2006).

Actualmente, la ciudad de Los Mochis cuenta con una superficie aproximada de 86.5 hectáreas, durante el censo de 2010 se registró una población de 256,623 habitantes (Dávila Ortiz, 2019; H. Ayuntamiento de Ahome, 2014), mientras que en el más reciente censo (2020) se contabilizaron 298,000 (INEGI, 2020) por lo que se puede hablar de un crecimiento demográfico del 16.12 % en los últimos 10 años.

Como se describió en esta sección, el desarrollo de Los Mochis, Sinaloa fue posible gracias a la conjunción de tres principales factores: ubicación estratégica, vías de comunicación y fácil acceso (bien se dice que Los Mochis es una de las pocas ciudades a las que se puede acceder por mar, aire, tren y carretera.; Dávila Ortiz, 2019), y la fertilidad de sus tierras debida principalmente a la creación de los canales de riego en una región de clima seco. Sin lugar a dudas, la fundación y eventual crecimiento de esta ciudad, está estrechamente vinculada con la red de riego del Valle del Río Fuerte, hoy conocido como Distrito de Riego 075 “Río Fuerte”.

3. Descripción del medio físico

3.1 Hidrografía

La ciudad de Los Mochis se encuentra dentro de la cuenca RH10Fb – B. Ohuira la cual pertenece a la cuenca Bahía Lechuguilla-Ohuira-Navachiste (RH10F), que a su vez forma parte de la Región Hidrográfica de Sinaloa (RH10; INEGI, 2010). La cuenca RH10Fb con un área de 2469.96 km^2 posee como corriente principal un escurrimiento intermitente, el cual surge en la parte alta de la cuenca y desemboca en la Bahía Ohuira (**Figura 1b**). En esta cuenca se originan varios escurrimientos de longitud restringida y de poca importancia que derivan su caudal hacia la bahía, cuerpos de agua internos o en alguna infraestructura de riego (Dávila Ortiz, 2019).

Hidrológicamente hablando, las mayores aportaciones de flujo provienen del Río Fuerte, importante afluente con origen en las estribaciones de la Sierra Tarahumara que drena sus aguas en los estados de Chihuahua y Sinaloa. El Río Fuerte cuenta con un sistema de presas (**Figura 1a**), a parir de las cuales deriva importantes volúmenes de agua para abastecer el Distrito de Riego 075, a través de una intrincada red de 2,297.79 km de canales sin revestir (H. Ayuntamiento de Ahome, 2014). Tanto el sistema de presas conformado por la Presas Huites, Josefa Ortiz y Miguel Hidalgo (**Figura 1a**) así como el Río Fuerte son de propiedad federal por lo cual son regulados por la Ley de Aguas Nacionales.

En cuanto a su hidrología subterránea, la ciudad de los Mochis se encuentra sobre el acuífero 2501 Río Fuerte, un acuífero libre que funge como zona de recarga entre la Sierra Madre Oriental y el Golfo de California (Dávila Ortiz, 2019).

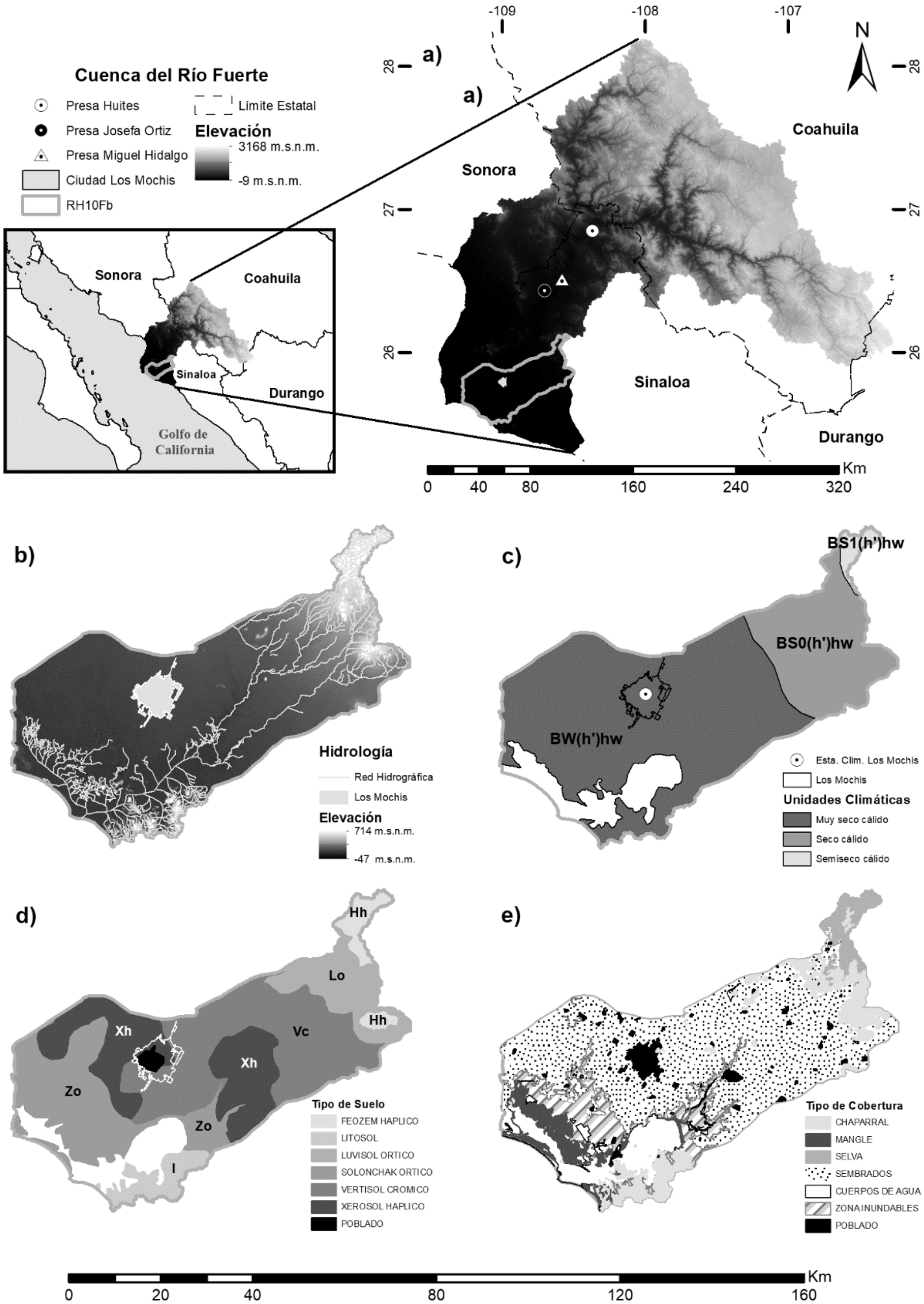


Figura 1. Caracterización de la zona de estudio con base en información del Instituto Nacional de Estadística, Geografía e Informática: Para ubicación de cuenca (a), hidrología (b), unidades climáticas según sistema de clasificación Köppen modificado por García (2004) (c), edafología (d) y tipo de uso de suelo y cobertura vegetal (e). Fuente: INEGI (2016).

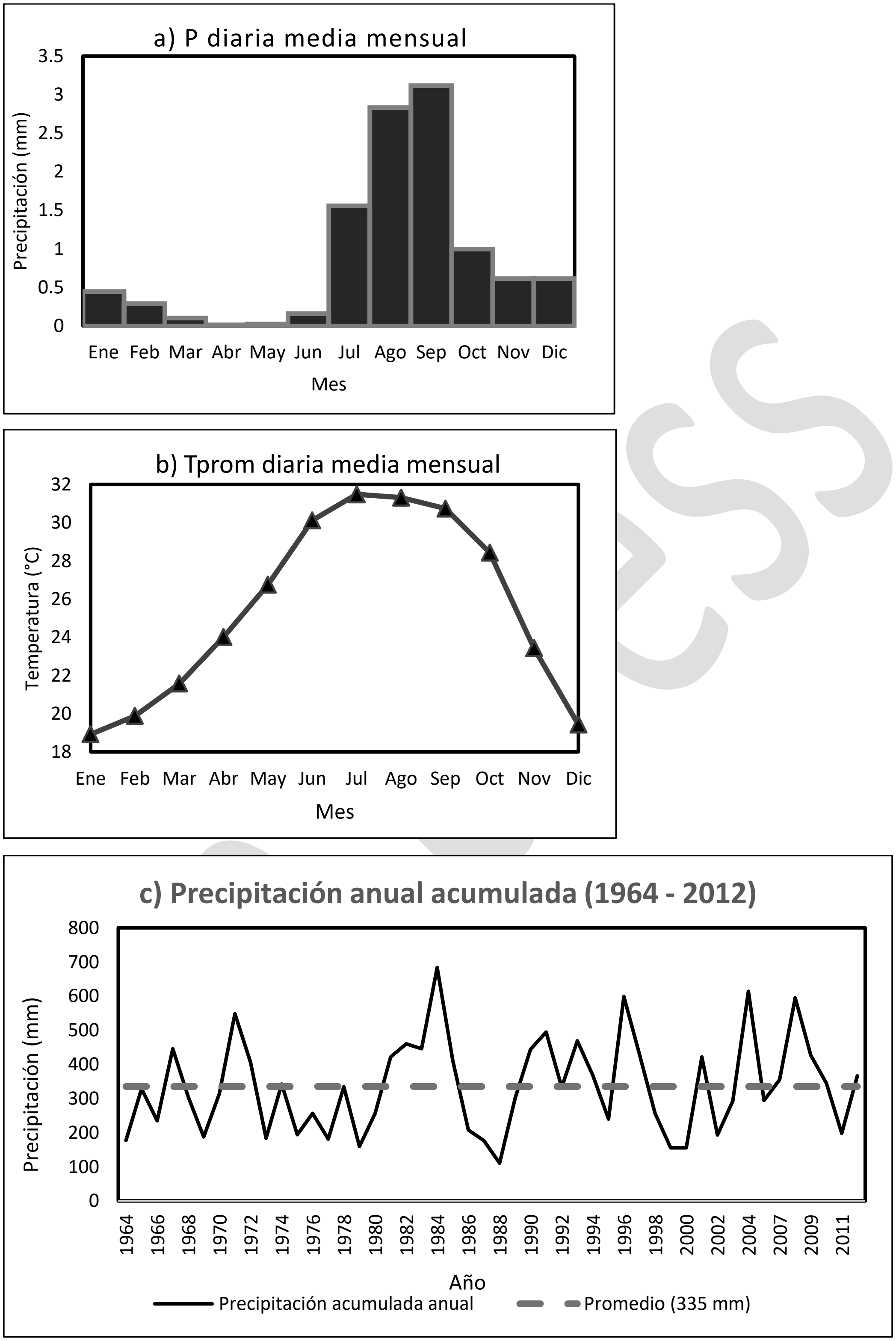


Figura 2. Precipitación (a), temperatura promedio (b) media mensual y precipitación anual (c) acumulada durante el periodo 1964-2012 estimados en la estación climática Los Mochis. Fuente: CLICOM (2016).

3.2 Clima

Según la clasificación climática de Köppen modificado por García (2004), el tipo de clima que prevalece en Los Mochis, es el Muy Seco Cálido (clave BW (h') hw), el cual predomina en la mayor parte de la cuenca RH10Fb (**Figura 1c**). El grupo de climas Secos (B) se caracterizan por precipitaciones anuales inferiores a los 800 mm, dentro de este grupo, el Clima Seco de Desierto (BW) es el más seco con precipitaciones inferiores a los 400 mm. Con respecto a sus características de temperatura, los climas cálidos (h') h, presentan una temperatura promedio anual se encuentra por encima de los 22° C.

En la **Figura 2** se muestran estimaciones de precipitación diaria media y de temperatura mensuales con base en datos recolectados por la estación Los Mochis (clave 265116; **Figura 1c**). Con respecto a la precipitación, se aprecia una marcada temporada de lluvias entre julio y septiembre, llegando hasta una máxima de 3 mm diarios, por su parte, en la temporada de secas la cual comienza en octubre y finaliza en junio se presentan varios meses con valores muy cercanos a 0 mm diarios. Este régimen de precipitación ha sido identificado en algunos reportes técnicos realizados en la zona, en los cuales se señala la ausencia de precipitación durante la mayor parte del año, concentrándose casi en su totalidad entre los meses julio y octubre, durante los cuales usualmente se presentan formaciones de tormentas y huracanes de gran intensidad (H. Ayuntamiento de Ahome, 2012; IMPLAN, 2012a).

Referente a los datos de precipitación anual acumulada registrada en la estación Los Mochis (**Figura 2c**), se ha estimado una precipitación anual acumulada promedio de 355 mm, por su parte, IMPLAN (2012) reporta 357.7 mm de lluvia anual acumulada entre 1981 y 2020. Observando los datos de precipitación anual acumulada durante el periodo 1964 y 2012, se puede apreciar una importante variabilidad en los años del registro. Comparando estos valores con eventos hidrometeorológicos extremos acontecidos en el área (**Tabla 2**), es posible identificar una clara relación entre los picos de precipitación en años lluviosos con eventos como huracanes, tormentas y depresiones tropicales o el fenómeno del Niño-oscilación del Sur, por ejemplo, la precipitación acumulada durante 1984 fue de 683.7 mm, año en el cual se registró la presencia de El Niño, para los años 1996 y 2008 se presentaron valores de precipitación de 599.3 y 595 mm (584 mm según IMPLAN, 2012) respectivamente, años en los que se reportaron los Huracanes Fausto y Olaf, los cuales generaron importantes afectaciones en la zona urbana de Los Mochis.

Analizando los datos de precipitación y de tipo de clima en la zona de estudio, se aprecia una interesante relación entre la presencia de eventos hidrometeorológicos severos con las condiciones climáticas de la zona, teniendo un régimen de lluvias que varía abruptamente mensual y anualmente en el cual prácticamente se concentra el mayor grueso de precipitación en solo tres meses. Aunado a esto, se aprecia el contraste entre el tipo de clima Muy Seco cálido BW (h') hw con este régimen de lluvias que podría catalogarse como extremo. Es por su régimen de lluvias y la constante presencia de eventos hidrometeorológicos extremos que se puede considerar a la precipitación como una condición física del medio altamente riesgosa.

Para el caso de la temperatura, en la **Figura 2** se puede apreciar una variación mensual entre los 18 y 32° C siendo junio, julio, agosto y septiembre los meses más cálidos y diciembre, enero y febrero los más fríos.

3.3 Edafología

Dentro de la cuenca RH10Fb se puede apreciar la presencia de cinco principales tipos de suelo (**Figura 1d**), Feozem háplico (Hh) suelo de profundidad variable de acuerdo a la topografía de terreno en la sección noreste de la cuenca donde se presentan las mayores elevaciones, posteriormente, se encuentran porciones de Luvisol órtico (Lo) suelo rojizo con un alto contenido en arcillas. En cuestión de extensión, el tipo de suelo predominante es Vertisol crómico (Vc), el cual se caracteriza por su alto contenido en arcillas expansivas las cuales en condiciones de saturación provocan que el terreno sea poco permeable (INEGI, 2004), su uso en agricultura es muy extenso debido a su alto nivel productivo, (gran parte del Distrito de Riego 075 presenta este tipo de suelo), no obstante, requiere de un buen drenaje debido al efecto de expansión-contracción de sus arcillas, el cual favorece el agrietamiento del terreno (INEGI, 2004). Otra unidad de suelo presente en la zona son los Vertisoles háplicos (Xh), tipo de suelo común en zonas áridas del centro y norte de México con bajo nivel de materia orgánica pero rico en arcillas (Ortiz-Villanueva & Ortiz-Solorio, 1990).

En la zona baja de la cuenca se presenta el tipo de suelo Solonchak órtico (Zo), el cual es común en zonas costeras, lechos de lagos y llanuras de inundación. Característico por su alto nivel de salinidad, este suelo puede ser inducido en zonas de riego por el ensalitramiento del terreno como resultado de un deficiente drenaje o un mal manejo del recurso hídrico (INEGI, 2004).

A nivel general se observa la presencia de suelos propios de zonas agrícolas, de ambientes áridos y zonas inundables, los cuales debido al riesgo de ensalitramiento del terreno o por la acción de su alto contenido de arcillas expansibles requieren de un buen sistema de drenaje para conservar sus niveles de producción agrícola. Esto último explica la presencia de sistema de los 1,634 km de drenes en todo el Distrito de Riego 075 (IMPLAN, 2012a) los cuales se encargan del desalojo excedentes del agua de riego en las zonas de cultivo.

3.4 Uso de suelo y vegetación

En la **Figura 1e**, se muestran los principales tipos de cobertura y uso de suelo presentes en la cuenca RH10Fb. De forma general, se aprecia la importante extensión que abarcan las zonas de cultivos del Distrito de Riego 075, las cuales rodean la zona urbana de Los Mochis y guardan una relación espacial con la distribución de Vertisol crómico en la cuenca (**Figura 1d**).

Con respecto al tipo de cobertura vegetal dentro de la cuenca, se aprecia una sucesión entre selva y chaparral en la parte alta de la cuenca, y presencia de vegetación de manglar en la zona costera. Así mismo, en el mapa de tipo de superficie, se muestra una significativa extensión de área con un alto nivel de susceptibilidad a inundarse. El anegamiento en estos sitos, se debe principalmente al ascenso de los niveles del mar, los cuales provocan la expansión de los límites de las Bahías Ohuira y Santa María. El suelo predominante en esa zona es el Solonchak órtico (Zo; **Figura 1e**), el cual se caracteriza por sus altas concentraciones de sales. La presencia de dicho suelo podría suponer una relación directa entre inundaciones costeras y el ensalitramiento del terreno (Dávila Ortiz, 2019).

3.5 Infraestructura hidráulica

3.5.1 Distrito de Riego 075 "Río Fuerte"

En secciones previas se ha hablado del Distrito de Riego 075, uno de los principales centros agrícolas del país (Pedroza González & Hinojosa Cuéllar, 2014), en el cual se cultivan principalmente maíz grano, papa y sorgo (CONAGUA & SEMARNAT, 2017). Este se encuentra en la porción norte de Sinaloa y comprende parte de los municipios de El Fuerte, Ahome, Guasave y Sinaloa, cuenta con una superficie con derecho a riego de 228,441 Ha y un total de 21,611 usuarios (Barrera Pérez, 2006).

El distrito de Riego 075 "Río Fuerte" está conformado por 13 módulos de riego y una Sociedad de Responsabilidad Limitada. De acuerdo a lo dispuesto en la Ley de Aguas Nacionales, mediante los Títulos de concesión de volúmenes, anualmente se otorgan 2,723.114 millones de m^3 de los cuales 2,623.114 millones de m^3 provienen del sistema de Presas Huites-Josefa Ortiz-Miguel Hidalgo mientras que 100 millones de m^3 son extraídos del acuífero del Río Fuerte (Barrera Pérez, 2006).

3.5.2 Red canales, drenes y colectores en Los Mochis

Con el fin de abastecer del recurso hídrico a los campos de cultivo presentes en el Distrito de Riego 075 se ha construido una vasta red de canales (2,297.79 km) de canales sin revestir; IMPLAN, 2012a) los cuales derivan agua proveniente del sistema de presas de la cuenca del Río Fuerte (**Figura 1**). Del mismo modo se cuenta con una intrincada red de drenes (1,634 km lineales de drenes; IMPLAN, 2012a). Este tipo de obras se excavan en tierra con un nivel topográfico menor a la de la superficie del terreno o en algunos casos se emplean arroyos naturales, y a diferencia de los canales, su función es desalojar excedentes de agua en zonas de cultivo ya sea excedentes del riego o de origen pluvial, en este caso, esto con el fin de evitar el ensalitramiento (El Debate, 2018c) o agrietamiento del suelo por el efecto de expansión-contracción por su alto contenido de arcillas.

A nivel local, una importante porción de la red de canales y drenes atraviesa la ciudad de Los Mochis. La presencia de este tipo de infraestructura atravesando la zona urbana es de vital importancia, ya que supone una (por no decir la más importante) de las condiciones de riesgo dentro de la zona urbana de la ciudad. Bajo este contexto, diversos medios de noticias locales (El Debate, 2018a, 2018b, 2018d, 2019c; López, 2018) y nacionales (Cabrera Martínez, 2018a), así como estudios técnicos en materia de análisis de riesgo realizados en el área de estudio (IMPLAN, 2012a; Organismo de Cuenca Pacifico Norte de la CONAGUA & SEMARNAT, 2016) y manifestaciones de impacto ambiental (H. Ayuntamiento de Ahome, 2012), concuerdan en que la deficiente e insuficiente infraestructura hidráulica superficial y subterránea constituye el factor de mayor peso que interviene en la generación de inundaciones y otros problemas de índole ambiental y urbana (IMPLAN, 2012b).

Esta problemática ha sido ampliamente documentada a través de medios informativos como sitios web y periódicos locales, en los cuales se hace énfasis en cómo, tanto la sociedad como las instituciones, identifican este tipo de infraestructura hidráulica como condiciones altamente riesgosas.

En la **Figura 3** se muestra el mapa de distribución de infraestructura hidráulica en Los Mochis conformada por drenes, canales, colectores pluviales y las plantas de tratamiento de aguas residuales, cada elemento de infraestructura hidráulica es identificado por un código numérico el cual se presenta en la **Tabla 1**. De particular importancia para las condiciones de riesgo en la ciudad se señalan al Dren Álamo (21) (López, 2018) y al Dren Juárez (22) (Cabrera Martínez, 2018a; El Debate, 2018a) como los que mayor impacto han tenido durante eventos de inundación.

En primera instancia, se cuenta con dos canales principales, el Canal Taxtes (8) y el Canal Lateral 18+420 (9) los cuales son derivaciones del Canal Principal Valle del Fuerte el cual proviene directamente del sistema de presas del Distrito de Riego 075, ambos canales recorren las periferias de la ciudad, conectándose a través del Canal Sublateral 23+700 (12) el cual resulta de vital importancia, ya que corta la ciudad de Oeste a Este y a partir de este, se derivan todos los canales y ramales los cuales dotan de agua a los ejidatarios de los poblados Benito Juárez, Plan de Ayala y 9 de diciembre, ubicados en el extremo sur de la ciudad (H. Ayuntamiento de Ahome, 2014). Además, parte de sus aguas alimentan la Planta Potabilizadora Ing. Terán Hernández (**Figura 3**), la cual posee una capacidad de producción de 800 l/s. La segunda planta potabilizadora conocida “Planta Mochis”, se localiza en las inmediaciones del Cerro de la Memoria, es alimentada por el Canal Hidalgo (9) y posee una capacidad de producción de 950 l/s (H. Ayuntamiento de Ahome, 2014).

Con respecto a la red de drenes, quizás el más importante en nivel de tamaño y caudal, es el Dren Juárez (22), el cual se ha convertido en un factor de riesgo hidrológico y ambiental para la ciudad de Los Mochis, ya que drena aguas excedentes agrícolas y pluviales generadas en aproximadamente 2,300 hectáreas de cultivo de la zona norte de la de la ciudad (H. Ayuntamiento de Ahome, 2012; Organismo de Cuenca Pacifico Norte de la CONAGUA & SEMARNAT, 2016), además, dentro del área urbana recibe importantes descargas provenientes del Dren Justicia Social (30) y del Dren Bay. 10 del Dren Juárez (23), y de los colectores pluviales Justicia Social (1) y Jiquilpan (2). En conjunto, estas causas provocan que en temporada de lluvias este dren se desborde provocando inundaciones en diversas zonas de la ciudad (Cabrera Martínez, 2018a; IMPLAN, 2012a), así mismo, se han reportado afectaciones ambientales y a la salud, como el ensalitramiento del suelo a causa del anegamiento de tierras de cultivo aledañas (El Debate, 2018c), o contaminación por tiraderos clandestinos instalados en los márgenes del dren (IMPLAN, 2012a).

En cuanto a infraestructura subterránea de ciudad Los Mochis, está la constituye una pequeña red de colectores pluviales en la cual, el agua de lluvia colectada se descarga en drenes a cielo abierto. Una situación que comúnmente se presenta en la zona urbana de Los Mochis durante eventos de inundación es el colapso de la red de colectores pluviales debido a la obstrucción de las rejillas de tormenta por residuos sólidos urbanos (Dávila Ortiz, 2019).

En un esfuerzo por mitigar los efectos adversos generados por los eventos de inundación, se han designado importantes recursos económicos para la restauración y ampliación de la red de drenaje urbano en la ciudad (El Heraldo de México, 2019).

Tabla 1. Listado de la red de canales, drenes y colectores pluviales que se encuentran en la ciudad de Los Mochis Sinaloa.

Red de Colectores Pluviales	Red de Canales	Red de Drenes
Blvd. Justicia Social (1)	Canal Taxtes (8)	Dren Álamo (21)
Pluvial Jiquilpan (2)	Canal Lateral 18+420 (9)	Dren Juárez (22)
Pluvial Blvd. Rosales (3)	Canal Hidalgo (10)	D Bay. 10 del Dren Juárez (23)
Pluvial Independencia (4)	Canal Ramal 0 (11)	Dren Mochicahui (24)
Mochis (Entubado) (5)	Canal Sublateral 23+700 (12)	Dren Cero (25)
Colector Río de las Cañas (6)	Canal Ramal 7+700 (13)	Dren Bay. 23+700 (26)
C. Pluvial Mochicahui Paralelo (7)	Ramal 6+700 (14)	Dren Mochis (27)
	Canal 5+700 (15)	Dren Miguelito (28)
	Ramal 4+700 (16)	Dren Ba. Flores Magón (29)
	Ramal 3+700 (17)	Dren Justicia Social (30)
	Canal Ramal 2+700 (18)	Dren Cañero (31)
	Canal Ramal 1+700 (19)	
	Canal Ramal 0 (20)	

Figura 3. Mapa de distribución de infraestructura hidráulica en Los Mochis en que muestra la red de colectores pluviales (líneas negras), canales principales (líneas compuestas gris y negro) y secundarios (lineras grises), drenes superficiales (líneas blancas) y drenes subterráneos (líneas blancas segmentadas), además, las principales plantas potabilizadoras. Fuente: elaboración propia con base en: IMPLAN (2012b, 2012c, 2012d).

Como dato importante, según IMPLAN (2012a) en esta urbanización no suele haber extracción de agua proveniente del subsuelo, ya que todo el recurso hídrico del que se dispone se obtiene de cuerpos de agua superficiales y de las plantas potabilizadoras.

4. Inundaciones en los Mochis

4.1 Inundaciones, una problemática histórica que prevalece

En Los Mochis los eventos de inundación representan una problemática histórica. Se han documentado inundaciones en la ciudad desde el año 1928 (aunque se tienen testimonios aún más antiguos; El Debate, 2018b), incluso, con el pasar de los años se puede apreciar que aparentemente esta tendencia va en aumento, en ese sentido, se tiene que tan solo en los últimos quince años se han presentado cuatro desastres por inundación en la ciudad, los cuales han generado severas afectaciones a la población, algunos llegando incluso a activar el Plan DN-III de la SEDENA (Plan de la Secretaría de la Defensa para el Auxilio a la Población Civil en Casos de Desastre; IMPLAN, 2012a).

En la **Tabla 2** se enlistan las inundaciones más importantes que se han presentado en el Municipio de Ahome, Sinaloa en los últimos 30 años, este recuento se realizó con base en información obtenida del Atlas de Riesgo de los Mochis (IMPLAN, 2012a) y de fuentes hemerográficas recopiladas por el Sistema de Inventario de Desastres DesInventar (Corporación OSSO & LA RED, 1994).

Tabla 2. Listado de las principales inundaciones que se ha presentado en el Municipio de Ahome, Sinaloa en los últimos 30 años. Elaborada con base en fuentes hemerográficas recopilados por el Sistema de Inventario de Desastres DesInventar (Corporación OSSO & LA RED, 1994) y el Atlas de Riesgo de los Mochis (IMPLAN, 2012a).

Fecha Inicio	Fuente	Observación de Efectos	Evento
01/10/1982	El Universal	Daños globales para todo el estado.	Huracán Paul
05/02/1983	Excélsior	Algunas hectáreas de jitomate y trigo están anegadas.	El Niño
04/10/1986	El Universal	Paralización de actividades. Daños infraestructura eléctrica	Huracán Paine
05/09/1995	CENAPRED, 1996	Precipitación con un máximo de 197 mm, afectaciones en diversas colonias	Lluvias Intensas
15/09/1996	El Universal	78 comunidades incomunicadas. 30 colonias inundadas. Lluvia Máxima acumulada 150 mm	Huracán Fausto
03/09/1998	La Jornada	Por lo menos 25 colonias inundadas. Aplican el Plan DN-III. Daños globales.	Huracán Isis
04/09/2004	La Jornada	Colapso de la Red de Drenaje en Los Mochis. Diez colonias damnificas y suspensión de las clases en nivel básico.	Lluvias
04/09/2007	SEDESOL, Gobierno Municipal, 2009	Evacuación de 260 familias en las partes bajas de los Mochis debido a lluvias. Precipitaciones pluviales de hasta 85 mm	Huracán Henriette

25/08/2008	Protección Civil	Colonias y calles afectadas por la lluvia. Suspensión de clases en diversos planteles educativos	Tormenta Tropical Junio
10/09/2008	La Jornada	Se suspendieron las clases en todos los niveles. El 90% de las colonias están inundadas. Municipio declarado en emergencia.	Tormenta Tropical Olaf
13/10/2009	La Jornada/Gobierno Municipal (2009)	Se aplica el Plan DN-III. Hasta un metro de altura. 600 personas trasladadas. Daños en viviendas, colapso de drenajes, cultivos inundados, calles y caminos vecinales. Los servicios de energía eléctrica y agua potable interrumpidos.	Tormenta Tropical Patricia
05/10/2014	LíneaDirectaPortal LD	Tromba deja inundaciones y tira árboles y espectaculares en Los Mochis y Guasave	Tromba
17/09/2018	El Universal/ Diversos medios	Lluvias provocan inundaciones en Los Mochis	Desborde Dren Juárez (22)

A pesar del alto grado de vulnerabilidad ante desastres por inundación que la ciudad de Los Mochis ha demostrado tener, actualmente existe un reducido número de trabajos de investigación en los que se aborde esta cuestión de una forma rigurosa, sistemática y científica, más allá de la información que se puede encontrar en estudios técnicos, como el Plan Operativo de Inundaciones (Organismo de Cuenca Pacifico Norte de la CONAGUA & SEMARNAT, 2016), manifestaciones de impacto ambiental (*e.g.*, H. Ayuntamiento de Ahome, 2012; Junta de Agua Potable y Alcantarillado del Municipio de Ahome, 2011; Toledo González, 2012), Programas Gubernamentales (*e.g.*, CONAGUA, 2012) y el Atlas de Riesgo Municipal de Ahome (IMPLAN, 2012a), siendo este último quizás el más importante, ya que en este trabajo se realizó un análisis de vulnerabilidad ante eventos de inundación, además se zonificaron las áreas más propensas a inundarse dentro de los Mochis sin embargo, no se especifican los criterios ni la metodología empleados para el mapeo de los polígonos de riesgo por inundación (Dávila Ortiz, 2019).

4.2 El caso de la inundación por la Depresión Tropical 19- E y la tardía intervención de las autoridades. "El desvío del Dren Juárez"

Durante el 17 y 21 de septiembre de 2018 se presentó la Depresión Tropical 19 - E, un evento hidrometeorológico extremo que tuvo lugar en diversas ciudades al noroeste del país. Las precipitaciones que se presentaron durante esos días generaron catastróficas inundaciones en diversas ciudades de Sinaloa, especialmente en Culiacán (El Sol de Sinaloa, 2019) y Ahome (Cabrera Martínez, 2018a, 2018b; El Debate, 2018b) en donde se reportaron afectaciones en un gran número de viviendas, campos de cultivo y el desplazamiento de miles de ciudadanos hacia albergues temporales (El Debate, 2018d).

Para el caso de la inundación acontecida en Los Mochis derivada de la Depresión Tropical 19-E, no existe un consenso general de la cantidad precisa de lluvia que se presentó, por ejemplo, el titular estatal de Protección Civil informó que las lluvias que se presentaron el 17 de septiembre, alcanzaron los 62 mm (Cabrera Martínez, 2018a), mientras que el coordinador de Protección Civil Municipal notificó 142 mm (Cabrera Martínez, 2018a). En ambos casos se habló de una "lluvia histórica" que no se había presentado desde hace diez años, no obstante, debido a que las autoridades consideraron que solo se habían presentado encharcamientos y no una inundación como tal, no se consideró activar el plan DN-III si no hasta el día siguiente, donde según El Debate (2018d) se presentaron lluvias que en total acumularon de 400 mm de agua, sin embargo, la nota no explica con claridad en que zona del municipio de Ahome se presentó la precipitación, ni la fuente de donde se obtuvo el dato. Posteriormente, medios locales reportaron un récord histórico de precipitación en la ciudad el día 19 de septiembre, en el cual según el coordinador de Protección Civil Municipal se presentaron 359.5 mm de lluvia de acuerdo a cifras de la CONAGUA dejando prácticamente damnificadas a todas las familias ahomenses (El Debate, 2018c), y generando importantes daños estructurales. De ser ciertas estas cifras, se estaría hablando de un evento de lluvia en el que precipitó la misma cantidad de agua que llueve durante un año en promedio. Sin embargo, registros climáticos en estaciones meteorológicas ubicadas en poblaciones aledañas, registraron una precipitación de 125.8 mm entre el 19 y 21 de septiembre (Dávila Ortiz, 2019).

El caso de inundación acontecida durante septiembre de 2018 en Los Mochis, es singular, ya que esta no es atribuida solamente al evento de precipitación extremo que se presentó en la zona. En ese sentido, diversos medios informativos e independencias como la presidencia municipal de Ahome y Protección Civil Municipal, identificaron como las principales causas de la anegación el desbordamiento de la mayoría de drenes y canales en la zona, el colapso de la red de colectores pluviales por la obstrucción de rejillas y bocas de tormenta, pero principalmente a la gran caudal de agua proveniente de los campos de cultivo aguas arriba de la zona urbana de Los Mochis a través del Dren Juárez (22) (Cabrera Martínez, 2018a, 2018b; El Debate, 2018b, 2018c, 2018d).

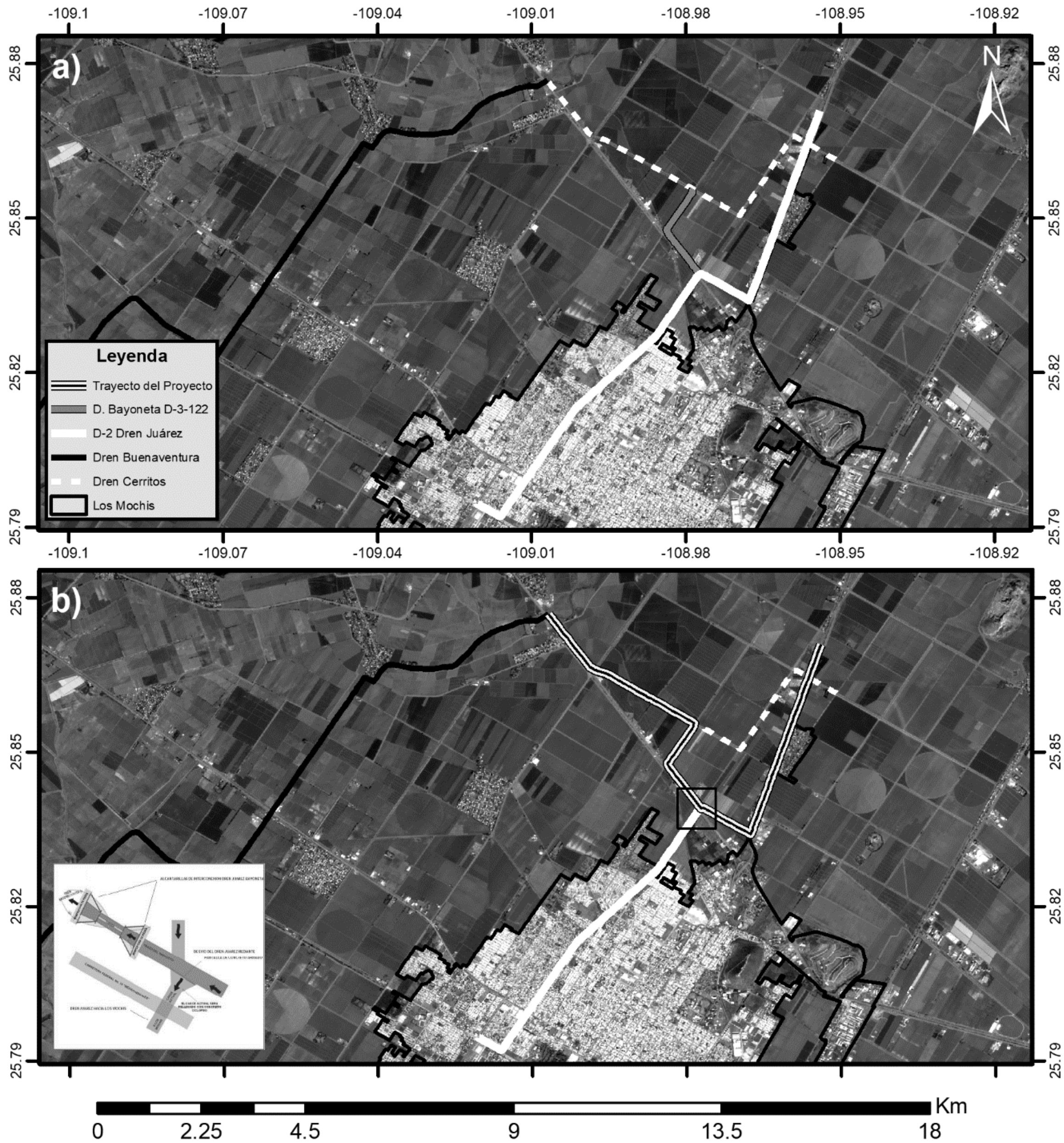


Figura 4. Estado actual del Dren Juárez (a) y la propuesta de proyecto para su desvío hacía el Dren Buenaventura a través de los drenes Bayoneta y Cerillos (b). El proyecto comienza con la obstrucción de flujo del Dren Juárez en el puente de la Carretera Federal No. 15 "México Nogales" y su interconexión con el Dren Bayoneta y termina con la descarga de agua residual de riego a 280 m de la línea de costa de Bahía Santa María. Fuente: H. Ayuntamiento de Ahome (2012); IMPLAN (2012c).

Previamente se ha hablado de la importancia del Dren Juárez (22) tanto como infraestructura hidráulica dentro de la red de drenaje del Distrito de Riego 075, como en su influencia dentro de las condiciones de riesgo de la zona urbana de Los Mochis. Anteriormente, se ha mencionado que el Dren Juárez recibe excedentes de riego de aproximadamente 2,300 ha de cultivo, aunando a eso, converge con el Dren Álamo (21) (otro dren del que se ha hablado anteriormente), justo en el centro de la de la urbe, a pocos metros del Canal Sublateral 23+700 (12) el cual transporta grandes volúmenes de agua entre extremos de la ciudad, creando así una potencial zona de inundación, la cual ya ha sido señala como tal en el más reciente Atlas de Riesgo Municipal (IMPLAN, 2012a).

Según El Debate (2019), el saldo final después de que se presentara la Depresión Tropical 19 – E en Sinaloa, fue de dos mil 900 personas evacuadas, once municipios declarados en zona de emergencia, perdidas en infraestructura, comunidades incomunicadas, pérdidas humanas, personas desaparecidas, etc. De acuerdo con el Director de Protección Civil de Sinaloa, se estimó que, en la ciudad de Los Mochis se tuvieron 70 mil viviendas dañadas y 230 mil personas damnificadas por las lluvias (Senado de la República, 2018).

Con motivo del grave problema de inundación que sufrió la ciudad de Los Mochis durante este evento hidrometeorológico, el Gobernador Quirino Ordaz junto con su esposa y el Alcalde de Ahome, Manuel Urquijo, recorrieron algunas de las zonas más afectadas por la inundación. Durante este recorrido, reconoció la importancia de llevar a cabo el proyecto de desviación del dren Juárez al dren Buenaventura, el cual se encontraba suspendido por cuestiones de financiamiento. En palabras del gobernador, “Es un proyecto que ya está en manos de CONAGUA y lo que ha faltado son los recursos porque la inversión total es arriba de 200 millones de pesos” (El Debate, 2018b) así mismo, el director de Protección Civil estatal, informo que se encontraban en proceso de elaboración de expediente para el Fondo de Desastres Naturales (FONDEN; El Debate, 2019).

Empezando por el proyecto “Conexión del dren Juárez al dren Buenaventura mediante la interconexión de tramos para la colecta de aguas pluviales”, este consiste en una propuesta para el desvió de este dren hacia el Dren Buenaventura, a través de los drenes Bayoneta y Cerillos ubicados en el límite norte de la ciudad (**Figura 4**). La idea de este proyecto, consiste en bloquear la entrada de flujo en el Dren Juárez a la altura de la carretera federal 15 “Navojoa - Los Mochis”, dirigiendo el agua excedente del distrito de riego aguas arriba, a través del Dren Buenaventura el cual desemboca sus aguas a 280 m de la línea de costa de Bahía Santa María. De acuerdo con estimaciones de la asociación civil “Por un Mochis mejor”, a través del desvío del dren Juárez se evitará que 5,000 ha de espejo de agua entren a la ciudad reduciendo en un 30% la probabilidad de presentar inundaciones (Bojórquez, 2020).

Un hecho importante, es que a pesar de que la manifestación de impacto ambiental de este proyecto (MIA-P 25SI2012HD044; H. Ayuntamiento de Ahome, 2012) fuera aprobada durante el año 2012, el director de la asociación “Por un Mochis mejor” señala que el proyecto lleva gestándose desde hace 35 años (Beltrán, 2021). Para 2015 solo se contaba con un temprano avance del kilómetro 0+500 al 3+300 que se realizó por parte del Gobierno del Estado de Sinaloa antes de quedar suspendido (Esthela García, 2020b), no fue hasta el día Martes 25 de septiembre de 2018 que el Sen. Mario Zamora Gastélum, exhortó a la Secretaría de Hacienda y Crédito Público a destinar de manera inmediata los recursos financieros para ejecutar este proyecto a través de una proposición con punto de acuerdo (Senado de la República, 2018).

Tal como se dispuso durante los recorridos del Gobernador a zonas afectadas por la inundación, más de un político se unió al llamado del Gobernador para solicitar el apoyo del Presidente Enrique Peña Nieto para declarar el estado de emergencia en Sinaloa a consecuencia de la Depresión tropical 19 – E y con la solicitud de apoyo ante el FONDEN. Con el cambio de mandato, el partido Morena en el Congreso Mexicano decidió cambiar de política en la prevención de riesgos y desastres, al aprobar en octubre de 2020 las modificaciones normativas que extinguen el FONDEN entre otros fideicomisos públicos (Añorve Baños, 2020). Sicairos (2019) en una incendiaria columna periodística, sugiere fallos sistémicos dentro del FONDEN, y habla de cómo el apoyo económico se ha proporcionado a cuentagotas pese al estado de emergencia de Sinaloa, además en dicha nota se insinúa un alto nivel de negligencia por parte de Protección Civil durante el siniestro de septiembre de 2018 en Los Mochis.

Finalmente, durante los últimos días de febrero de 2021, el director regional de la CONAGUA informó que a más tardar en 10 días quedarían concluidos los trabajos de la obra de desviación del Dren Juárez al Buenaventura (Beltrán, 2021). El proyecto se realizó en tres fases a través de procesos de licitación coordinados por la CONAGUA, La consolidación de este proyecto supuso un esfuerzo de 35 años de gestión y un presupuesto de 200 mdp (Esthela García, 2020b).

4.3 Afectaciones del dren Juárez, "pasándole el problema al vecino"

Esthela García (2020a) titula una nota como "Desviación del dren Juárez no representa riesgo para nadie: CONAGUA" en la cual se puntualiza la optimista postura tanto de la CONAGUA como del Presidente de la Cámara Mexicana de la Industria de la Construcción en Los Mochis, acerca el proyecto de desvío del Dren Juárez a través del Dren Buenaventura, el cual en ningún momento estuvo libre de controversias tanto por el manejo de apoyos económicos (Beltrán, 2021; Sicairos, 2019) como por sus implicaciones en el medio ambiente (Beltrán, 2019; El Debate, 2019b, 2019a). Lo cierto es que tal declaración emitida por la CONAGUA no es del todo cierta, ya que a la fecha se han reportado ciertas cuestiones no consideradas durante la propuesta y construcción de esta obra hidráulica. En primer lugar, 2 km antes de llegar a la desembocadura del Dren Buenaventura en la bahía Santa María a la altura del ejido Bachomobampo el cauce de las aguas que corren por el Dren Juárez se encuentran obstruidas y reembalsadas por un taponamiento de 2.3 km de manglar (El Debate, 2019a). La distribución de manglar en la zona costera de la cuenca RH10Fb, se muestra en la **Figura 1e**. Ante esta situación, el presidente del comisariado ejidal del ejido Tortugas 2, junto con otros ejidatarios han manifestado que este es un problema al "que se le ha estado dando vueltas" durante mucho tiempo, ya que no se cuenta con un permiso de la SEMARNAT para el desfogue del dren (El Debate, 2019b), cabe mencionar las cuatro especies de manglar presentes en México están sujetas a protección especial de acuerdo a la NOM 059 SEMARNAT-2010 (DOF, 2010).

Hasta este punto, se cuenta con un ambicioso proyecto de infraestructura hidráulico para evitar el acaecimiento de inundaciones, el cual se encuentra inhabilitado debido a que esta obstruido por una zona de manglar, esto nos lleva al segundo impacto. Previamente se ha hablado de la importancia del drenaje de suelo debido a las propiedades edafológicas del terreno. Los suelos salinos, como los presentes en esta zona, generan fuertes afectaciones en la productividad agrícola debido al ensalitramiento cuando se cuenta con drenaje insuficiente. En ese sentido, el presidente de comisariado ejidal del ejido Mochis 2, mencionó que se tienen 222 ha afectadas por altos niveles de salinidad y que se estima que entre 600 y 800 hectáreas presentan afectaciones en mayor o menor medida (El Debate, 2019a). Por su parte, el ex comisariado ejidal de Plan de Guadalupe, señalo que los desbordamientos del Dren Juárez, ocasionados por el taponamiento en su desembocadura con el mar, dejaron pérdidas millonarias en terrenos agrícolas, ganado y viviendas en dicha comunidad (Beltrán, 2019). Ante este panorama, se ha puesto sobre la mesa la importancia de limpiar, desazolvar y rehabilitar drenes y colectores pluviales antes que el desvío del Dren Juárez, mientras tanto, la problemática prevalece.

El caso del desvío del Dren Juárez es un ejemplo más de como la falta de un profundo conocimiento sobre los elementos que conforman el espacio y sus interacciones naturales y sociales, conllevan a la perpetuación de las condiciones de vulnerabilidad ante eventos perturbadores como consecuencia de una mala planeación. El proyecto de desviación del Dren Juárez ha sido un tema recurrente ante las catastróficas inundaciones que se han presentado en Los Mochis Sinaloa, una obra hidráulica que transporta grandes volúmenes de agua a través de una ciudad por la cual no corre ningún río, arroyo o meandro de forma natural, en una cuenca de escurrimiento intermítete, todo en aras del desarrollo agrícola del país, en el cual se puede apreciar el rol fundamental que juegan las condiciones físicas del medio, el ordenamiento territorial, la planeación de infraestructura hidráulica y las políticas públicas.

5. Conclusiones

Los Mochis es una ciudad joven que fue fundada a finales del siglo XIX con la idea de desarrollar una región con potencial de recursos hídricos (provenientes del caudal del Río Fuerte) y convertirla así en una región fértil para la agricultura y propicia para el comercio. De esta manera la ciudad creció entre los canales de riego y los drenes de desagüe del excedente de agua de los campos de cultivo. La ciudad se desarrolló en el contexto de la creación del distrito de riego 075 y la construcción de embalses para el suministro de agua a dicho sistema. De esta manera, la población de Los Mochis se ha encontrado en un entorno donde las inundaciones, provocadas por el desbordamiento de la red de canales y drenes, son frecuentes. Los eventos de inundación se han documentado desde 1918 hasta la actualidad y muchos han tenido consecuencias graves, entre estos destaca la anegación del 90% de las colonias de la ciudad en 2008. Estos eventos generalmente están asociados a un evento hidrometeorológico en específico, como tormentas tropicales, huracanes y la presencia del fenómeno de El Niño.

Otro evento importante es la inundación de septiembre de 2018 que causó afectaciones a un gran número de viviendas, campos de cultivo y la evacuación de miles de personas. Se atribuyó esta inundación, por parte de las autoridades, a una "lluvia histórica", sin embargo, hay diversos datos contradictorios sobre la intensidad de esta precipitación que va de 400 mm reportados en la prensa a 135 mm consultados en los registros de estaciones meteorológicas cercanas. Lo que es claro es que este evento de inundación y las afectaciones no son atribuibles directamente a la precipitación sino al desbordamiento de la red de canales y drenes de la ciudad.

Las inundaciones históricas de la ciudad están condicionadas por la conformación del espacio en el que se ha desarrollado Los Mochis. La ciudad ha sido construida en un contexto hidrológico muy particular, pues en ella no corren ríos ni arroyos naturales y está localizada una región con clima seco. La progresión del riesgo ante eventos de inundación se ha incrementado por diversos factores, entre los que destaca el crecimiento de la población y del área de urbanización, provocando, según la percepción de los pobladores, un aumento en la frecuencia de los eventos severos de inundación. En ese sentido, se tiene que tan solo en los últimos quince años se han presentado cuatro desastres por inundación en la ciudad, los cuales han generado severas afectaciones a la población, algunos han requerido la activación del Plan DN-III de la SEDENA.

La solución al problema de las inundaciones en Los Mochis ha tenido como base principal una propuesta técnica para el desvío del dren Juárez hacia el dren Buenaventura, el cual reduciría en 30% la probabilidad de inundaciones en la ciudad, según los estudios. El proyecto lleva gestándose 35 años. Sin embargo, esta propuesta no está exenta de controversia pues podría afectar una zona de manglares en la desembocadura del dren Buenaventura.

Este estudio de caso nos permite interpretar la progresión del riesgo ante inundaciones como una conformación de factores han propiciado el incremento de la vulnerabilidad de la población a través del tiempo; entre estos, la precipitación tiene un peso menos determinante que el contexto histórico de conformación de la ciudad dentro de un sistema de riego.

6. Referencias

Añorve Baños, M. (2020, noviembre 20). El Fonden no debió haber desaparecido. *El Universal*. Recuperado de https://www.eluniversal.com.mx/opinion/manuel-anorve-banos/el-fonden-no-debio-haber-desaparecido

Arreguín-Cortés, F. I., López-Pérez, M., Marengo-Mogollón, H., Arreguín-Cortés, F. I., López-Pérez, M., & Marengo-Mogollón, H. (2016). Las inundaciones en un marco de incertidumbre climática. *Tecnología y ciencias del agua*, *7*(5), 5–13.

Barrera Pérez, J. J. (2006). *Análisis del Distrito de Riego 075, El Fuerte, Sinaloa, México* (Tesis para obtener el grado de Maestría en Ingeniería Civil - Aprovechamientos Hidráulicos, Universidad Autónoma de México). Universidad Autónoma de México, México. Recuperado de http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/676/barreraperez.pdf?sequence=10

Beltrán, K. (2019, junio 27). Desviación Dren Juárez, no solucionará inundaciones. *Meganoticias*. Recuperado de https://www.meganoticias.mx/los-mochis/noticia/desviacion-dren-juarez-no-solucionara-inundaciones/82511

Beltrán, K. (2021, febrero 26). “Un logro de la sociedad civil”, obra desviación Dren Juárez. *Meganoticias*. Recuperado de https://www.meganoticias.mx/los-mochis/noticia/un-logro-de-la-sociedad-civil-obra-desviacion-dren-juarez/223582

Blaikie, P., Cannon, T., Davis, I., & Wisner, B. (1996). *Vulnerabilidad: El Entorno Social, Político y Económico de los Desastres* (Primera Edición). Colombia: Red de Estudios Sociales en Prevención de Desastres en América Latina. Recuperado de https://www.desenredando.org/public/libros/1996/vesped/vesped-todo_sep-09-2002.pdf

Bojórquez, C. (2020, junio 30). Lanzan licitación para segunda etapa de la desviación del dren Juárez en Los Mochis. *NORESTE*. Recuperado de https://www.noroeste.com.mx/amp/norte/lanzan-licitacion-para-segunda-etapa-de-la-desviacion-del-dren-juarez-en-los-mochis-LYNO1200496

Cabrera Martínez, J. (2018a, septiembre 19). Fuertes lluvias provocan inundaciones en Los Mochis, Sinaloa; suspenden clases. *El Universal*. Recuperado de https://www.eluniversal.com.mx/estados/fuertes-lluvias-provocan-inundaciones-en-los-mochis-sinaloa-suspenden-clases

Cabrera Martínez, J. (2018b, septiembre 20). Lluvias en Sinaloa ocasionan inundaciones y accidentes. *El Universal*. Recuperado de http://www.eluniversal.com.mx/estados/lluvias-en-sinaloa-ocasionan-inundaciones-y-accidentes-reportan

CLICOM. (2016). *Base de Datos Climatológica Nacional. Servicio Meteorológico Nacional a través del CICESE*. [Base de datos]. Recuperado de http://clicom-mex.cicese.mx.

CONAGUA. (2012). *Programa Hídrico Regional Visión 2030. Región Hidrológico-Administrativa III Pacífico Norte* (SEMARNAT, Ed.). México, D.F.: Coordinación General de Atención Institucional, Comunicación y Cultura del Agua de la Comisión Nacional del Agua. Recuperado de http://www.conagua.gob.mx/conagua07/publicaciones/publicaciones/3-sgp-17-12pn.pdf

CONAGUA & SEMARNAT. (2017). *Estadísticas Agrícolas de los Distritos de Riego. Año Agrícola 2015-2016*. Ciudad de México, México. Recuperado de https://files.conagua.gob.mx/conagua/publicaciones/Publicaciones/EA_2015-2016.pdf

Corporación OSSO & LA RED. (1994). *Sistema de Inventario de Desastres. DesInventar*. [Base de datos]. Recuperado de https://www.desinventar.org/es/

Dávila Ortiz, R. (2019). *Respuesta hidrológica en zonas urbanas ante eventos severos de tormenta aplicando modelación numérica e información SAR: Caso Los Mochis*. Recuperado de https://repositorio.ipicyt.edu.mx///handle/11627/5259

DOF. *NORMA Oficial Mexicana NOM-059-SEMARNAT-2010, Protección ambiental-Especies nativas de México de flora y fauna silvestres-Categorías de riesgo y especificaciones para su inclusión, exclusión o cambio-Lista de especies en riesgo.* , (2010).

El Debate. (2018a, septiembre 17). Los Mochis del ayer, no es muy distinto al de ahora. *El Debate*. Recuperado de https://www.debate.com.mx/losmochis/los-mochis-inundaciones-historia-calles-lluvia-20180917-0151.html

El Debate. (2018b, septiembre 18). 90 mm de lluvia y 12 toneladas de basura; el recuento de daños. *El Debate*. Recuperado de https://www.debate.com.mx/losmochis/fuertes-lluvias-mochis-sinaloa-inundaciones-danos-20180918-0013.html

El Debate. (2018c, septiembre 20). Los Mochis, Sinaloa bajo el agua, estos son los estragos de las fuertes lluvias. *El Debate*. Recuperado de https://www.debate.com.mx/losmochis/lluvias-inundaciones-sinaloa-los-mochis-el-fuerte-clima-mexico-20180920-0011.html

El Debate. (2018d, septiembre 21). Depresión 19-E baja su fuerza pero deja a Ahome bajo el agua. *El Debate*. Recuperado de https://www.debate.com.mx/losmochis/los-mochis-sinaloa-inundaciones-depresion-tropical-lluvias-20180921-0021.html

El Debate. (2019a, agosto 24). El dren Juárez, uno de los riesgos de inundación en Los Mochis. *El Debate*. Recuperado de https://www.debate.com.mx/losmochis/El-dren-Juarez-uno-de-los-riesgos-de-inundacion-en-Los-Mochis-20190824-0115.html

El Debate. (2019b, agosto 24). Se desborda el dren Juárez por taponamiento a la salida al mar. *El Debate*. Recuperado de https://www.debate.com.mx/losmochis/Se-desborda-dren-Juarez-por-taponamiento-a-la-salida-al-mar-20190825-0009.html

El Debate, J. (2019c, septiembre 17). A un año que Los Mochis quedó bajo el agua. *El Debate*. Recuperado de https://www.debate.com.mx/losmochis/A-un-ano-que-Los-Mochis-quedo-bajo-el-agua-20190917-0121.html

El Heraldo de México. (2019, febrero 28). Invierten 120 mdp en drenaje Rehabilitarán colectores de agua en Los Mochis para desalojar aguas pluviales. *El Heraldo de México*. Recuperado de https://heraldodemexico.com.mx/estados/invierten-120-mdp-en-drenaje/

El Sol de Sinaloa. (2019, septiembre 19). *A un año de la Depresión Tropical 19-E, la ayuda que nunca llegó...* Recuperado de https://www.elsoldesinaloa.com.mx/local/a-un-ano-de-la-depresion-tropical-19-e-la-ayuda-que-nunca-llego...-4199355.html

García, Enriqueta. (2004). *Modificaciones al Sistema de Clasificación Climática de Köppen* (Quinta Edición). México: UNAM. Instituto de Geografía. Recuperado de http://www.igeograf.unam.mx/sigg/utilidades/docs/pdfs/publicaciones/geo_siglo21/serie_lib/modific_al_sis.pdf

García, Esthela. (2020a, septiembre 7). Inicia la segunda etapa de desviación del dren Juárez en Los Mochis. *Luz Noticias*. Recuperado de https://www.luznoticias.mx/2020-09-07/sinaloa/inicia-la-segunda-etapa-de-desviacion-del-dren-juarez-en-los-mochis/98863

García, Esthela. (2020b, noviembre 20). Desviación del dren Juárez no representa riesgo para nadie: Conagua. *Luz Noticias*. Recuperado de https://www.luznoticias.mx/2020-11-20/sinaloa/desviacion-del-dren-juarez-no-representa-riesgo-para-nadie-conagua/103844

Golnaraghi, M., Etienne, C., Guha-Sapir, D., & Below, R. (2014). *Atlas of Mortality and Economic Losses from Weather, Climate, and Water Extremes (1970-2012)* (M. Williams & C. Sylvie, Eds.). Geneva 2, Switzerland: World Meteorological Organization. Recuperado de https://library.wmo.int/pmb_ged/wmo_1123_en.pdf

H. Ayuntamiento de Ahome. (2012). *Manifestación de Impacto Ambiental. Modalidad Particular, Sector Hidráulico (25SI2012HD044) – Conexión del Dren Juárez al Dren Buenaventura y rehabilitación mediante la interconexión de los tramos 0+000 al 2+445 (Dren D3-122) y del tramo 0+000 al 5+400 (Dren D-3 168) para la colecta de aguas pluviales y de excedentes agrícolas como obras de protección para la ciudad de Los Mochis, Ahome, Sinaloa.* Ahome, Sinaloa. Recuperado de http://sinat.semarnat.gob.mx/dgiraDocs/documentos/sin/estudios/2012/25SI2012HD044.pdf

H. Ayuntamiento de Ahome. (2014). *Actualización del Plan Director de Desarrollo Urbano de la Ciudad de Los Mochis*. Recuperado de http://www.implanahome.gob.mx/pddu_los_mochis_2014.html

H. Ayuntamiento de Ahome. (2018). Historia—Reseña Histórica. Recuperado el 19 de marzo de 2021, de Transformando Ahome Escribiendo una nueva historia website: https://www.ahome.gob.mx/historia/

IMPLAN. (2012a). *Atlas de Riesgos Naturales de la ciudad de Los Mochis 2011* (p. 109). Ahome, Sinaloa: Instituto Municipal de Planeación de Ahome. Recuperado de Instituto Municipal de Planeación de Ahome website: http://rmgir.proyectomesoamerica.org/PDFMunicipales/2011/vr_25001_AR_AHOME.pdf

IMPLAN. (2012b). *MAPA 01.- RED DE CANALES* [[Plano Cartográfico]]. En *Mapas Técnicos. Los Mochis*. Los Mochis, Ahome, Sinaloa: H. Ayuntamiento de Ahome. Recuperado de http://www.implanahome.gob.mx/mapas_tecnicos.html

IMPLAN. (2012c). *MAPA 02.- RED DE DRENES* [[Plano Cartográfico]]. En *Mapas Técnicos. Los Mochis*. Los Mochis, Ahome, Sinaloa: H. Ayuntamiento de Ahome. Recuperado de http://www.implanahome.gob.mx/mapas_tecnicos.html

IMPLAN. (2012d). *MAPA 03.- RED DE COLECTORES* [[Plano Cartográfico]]. En *Mapas Técnicos. Los Mochis*. Los Mochis, Ahome, Sinaloa: H. Ayuntamiento de Ahome. Recuperado de http://www.implanahome.gob.mx/mapas_tecnicos.html

INEGI. (2004). *Guías para la Interpretación de Cartografía. Edafología*. Aguascalientes, México: Instituto Nacional de Estadística, Geografía e Informática. Recuperado de www.inegi.org.mx/inegi/SPC/doc/INTERNET/EDAFI.pdf

INEGI. (2010). *Red Hidrográfica escala 1:50 000* [[Formato Shapefile]]. En *Cartas Climatológicas* (2° Edición). Instituto Nacional de Estadística, Geografía e Informática. Recuperado de https://www.inegi.org.mx/temas/mapas/hidrografia/

INEGI. (2016). *Datos en formato vectorial de México* [[Formato Shapefile]]. Instituto Nacional de Estadística Geografía e Informática. Recuperado de http://www.inegi.org.mx/geo/contenidos/recnat/default.aspx

INEGI. (2020). *Censo de Población y Vivienda 2020*. México. Recuperado de https://www.inegi.org.mx/programas/ccpv/2020/

Infante, J. A. (2011, junio 6). La verdadera historia del origen de Los Mochis. *Globedia*. Recuperado de http://mx.globedia.com/verdadera-historia-origen-mochis

Instituto Latinoamericano de la Comunicación Educativa. (2021). Los Mochis y el valle del Fuerte. Recuperado de Biblioteca Digital del ILCE website: http://bibliotecadigital.ilce.edu.mx/sites/estados/libros/sinaloa/html/sec_100.html

Junta de Agua Potable y Alcantarillado del Municipio de Ahome. (2011). *Manifestación de Impacto Ambiental. Modalidad Particular (25SI2011HD100)—"Construcción, Operación y Mantenimiento" del Colector Oriente (segunda etapa), ubicado en la Ciudad de Los Mochis, Ahome, Sinaloa.* Los Mochis, Ahome, Sinaloa. Recuperado de http://sinat.semarnat.gob.mx/dgiraDocs/documentos/sin/estudios/2011/25SI2011HD100.pdf

Ley García, J., & Calderón Aragón, G. (2008). De la vulnerabilidad a la producción del riesgo en las tres primeras décadas de la ciudad de Mexicali, 1903-1933. *Región y sociedad*, *20*(41), 145–173.

López, A. (2018, octubre 27). Dren Álamo es el más dañino para la ciudad de Los Mochis. *El Debate*. Recuperado de https://www.debate.com.mx/sinaloa/losmochis/dren-alamo-mas-peligroso-los-mochis-estudio-20181027-0006.html

mochisonline. (2019, junio 18). Los Mochis a sus 109 años. *mochisonline.com*. Recuperado de http://www.mochisonline.com/170612los_mochis_109_aniversario.html

Organismo de Cuenca Pacifico Norte de la CONAGUA & SEMARNAT. (2016). *Plan Operativo de Inundaciones de la Ciudad de Los Mochis, Sinaloa.* Culiacán, Sinaloa, México: Coordinación General de Atención de Emergencias y Consejos de Cuenca Organismo de Cuenca Pacífico Norte. Recuperado de https://mapasin.files.wordpress.com/2017/07/conagua-plan-operativo-de-inundaciones-los-mochis-2017.pdf

Ortiz-Villanueva, B., & Ortiz-Solorio, C. (1990). *Edafología*. México: Universidad Autónoma Chapingo, Departamentos de Suelos.

Pedroza González, E., & Hinojosa Cuéllar, G. (2014). *Manejo y distribución del agua en distritos de riego: Breve introducción didáctica* (2° Edición). Juitepec, Morelos, México: Instituto Mexicano de Tecnología del Agua. Recuperado de https://www.imta.gob.mx/biblioteca/libros_html/manejo-dadr/files/assets/common/downloads/publication.pdf

Quintana Navarrete, J. (2015, octubre 20). La utopía está en otra parte: Albert K. Owen y la colonización de Topolobampo. *Horizontal*. Recuperado de https://horizontal.mx/la-utopia-y-topolobampo/

Rodríguez Esteves, J. M. (2007). La conformación de los "desastres naturales": Construcción social del riesgo y variabilidad climática en Tijuana, B. C. *Frontera norte*, *19*(37), 83–112.

Senado de la República. (2018, septiembre 27). *Solicitud para la resolución de la: Proposición con punto de acuerdo por el que el Senado de la República exhorta la Secretaría de Hacienda y Crédito Público destine de manera inmediata los recursos financieros para ejecutar el proyecto denominado "Conexión de Dren Juárez al Dren Buenaventura mediante la interconexión de tramos para la colecta de aguas* *pluviales”.* Recuperado de http://www.senado.gob.mx/64/gaceta_del_senado/documento/83605

Sicairos, A. (2019, agosto 25). El Fonden le ha mentido a Sinaloa. *NORESTE*. Recuperado de https://www.noroeste.com.mx/colaboraciones/el-fonden-le-ha-mentido-a-sinaloa-DGOP110074

Toledo González, J. F. (2012). *Manifestación de Impacto Ambiental. Modalidad Particular (25SI2012HD028)—Entubamiento del dren colector pluvial Mochis, tramo del km 5+724.75 al km 6+763.85, y construcción de estructuras complementarias*. Ahome, Sinaloa. Recuperado de http://sinat.semarnat.gob.mx/dgiraDocs/documentos/sin/estudios/2012/25SI2012HD028.pdf

Cazando nubes de tormenta. Propuesta para un sistema de monitoreo y alerta temprana basado en inteligencia artificial en México.

enero 11, 2024

MC Rodrigo Dávila Ortiz y Dr. José Tuxpan Vargas. Laboratorio de Geomática y Modelación Numérica. División de Geociencias Aplicadas. Instituto Potosino de Investigación Científica y Tecnológica, A.C. (IPICYT).

Si alguna vez observaste el cielo momentos previos a que se desatara una tormenta, posiblemente hayas sido testigo de la formación de una nube de convección profunda o también conocidas como Cumulonimbus. Estas majestuosas nubes se caracterizan por su gran tamaño y desarrollo vertical, con una cima aplanada la cual recuerda la forma de un yunque y debido a su gran contenido de agua, estas pueden llegar a bloquear la luz del sol presentando una base obscura, dándole al cielo su peculiar color gris durante estos eventos.

Las nubes de tormenta se generan a partir de un proceso de convección profunda, donde una masa de aire cálido y húmedo asciende de manera intensa a través de la atmósfera en forma de espiral rotatoria. Aunque existen diferentes mecanismos de formación de estas nubes, como el borde de un frente frío, el choque con otras masas de aire (convergencia en superficie) o con una cadena montañosa (levantamiento orográfico), estas se caracterizan por dos cosas, la rapidez con la que pueden llegar a formarse y el estar asociadas con condiciones meteorológicas extremas, como tormentas eléctricas, lluvias intensas, caída de granizo, fuertes vientos y, en algunos casos, tornados. Simplemente, durante 2020 se estima que la presencia de tormentas causó la muerte de más de 1,700 personas en todo el mundo.

Por su parte, México tiene un alto grado de vulnerabilidad ante eventos hidrometeorológicos extremos, debido a su complejo relieve, su posición geográfica, estar rodeado de océanos, pero especialmente por sus condiciones socio-económicas. Recientes inclemencias meteorológicas a lo largo del país han puesto de manifiesto la importancia de comprender y abordar los eventos de convección profunda, así como los fenómenos atmosféricos que pueden desencadenar condiciones climáticas extremas y representar riesgos significativos para la sociedad.

En este contexto, en el Laboratorio de Geomática y Modelación Numérica de la División de Geociencias Aplicadas del IPICYT, estamos trabajando en el desarrollo de un sistema de monitoreo y alerta temprana de inundaciones asociadas a la presencia de nubes de tormenta, bajo tres principios clave, 1) el uso de productos de acceso libre y software de código abierto, lo cual permite que la implementación de este sistema sea asequible, 2) que sea automático, desde su diseño, se ha planteado que este sistema sea autónomo y pueda emitir alertas en tiempo real y, 3) que sea escalable para cualquier zona de México.

Esta metodología se basa en el uso de modelos de Inteligencia Artificial, estos, al ser entrenados con eventos de referencia, aprenden patrones con los que pueden hacer predicciones con datos nuevos. Para alimentar estos modelos, se utilizan datos satelitales de una red de satélites llamados GOES-R operada por la Administración Nacional Oceánica y Atmosférica (NOAA) de los Estados Unidos y diseñada para proporcionar observaciones meteorológicas continuas desde la órbita geoestacionaria, esto último significa que los satélites se mueven de acuerdo al periodo de rotación de la tierra, lo cual les permite registrar el mismo punto siempre. Los satélites GOES-R tienen la capacidad de monitorear fenómenos meteorológicos en tiempo real (cada 5 minutos para México), lo que incluye observaciones de nubes, mapeo de rayos, incendios forestales, actividad solar y otros eventos climáticos.

Actualmente, se ha probado esta metodología de forma exitosa en Los Mochis, Sinaloa y la Ciudad de México (CDMX), en las siguientes etapas de este proyecto se abordará la integración de este sistema de detección de nubes de tormenta con datos de la red estaciones climáticas de México y con un modelo hidrológico para la generación de mapas de potencial riesgo ante inundaciones con un periodo de actualización cada 5 minutos.

Con esta investigación buscamos prevenir y mitigar los riesgos asociados con eventos de tormenta al proporcionar información oportuna y precisa a tomadores de decisión y potenciales personas afectadas, además de mejorar nuestro entendimiento sobre la formación y evolución de este tipo de fenómenos en nuestro país.